Article:

Sweeney, Rohan, Suhrcke, Marc Eckart orcid.org/0000-0001-7263-8626, Jeon, Yun Joo et al. (1 more author) (2018) The impact of SWAps on health aid displacement of domestic health expenditure. JOURNAL OF DEVELOPMENT STUDIES. pp. 719-737. ISSN 0022-0388

https://doi.org/10.1080/00220388.2017.1303670

Title: The impact of SWAps on health aid displacement of domestic health expenditure.

Rohan Sweeney[1,2‡]
Marc Suhrcke[3]
Yun Joo Jeon[4]
Duncan Mortimer[2]

[1] Deakin Health Economics, Deakin University, Geelong, Australia
[2] Centre for Health Economics, Monash Business School, Monash University, Melbourne, Australia
[3] Centre for Health Economics, University of York, York, UK.
[4] Agency for Care Effectiveness, Ministry of Health, Singapore
‡ Corresponding author. Address: Deakin Health Economics, Deakin University. 221 Burwood Highway, Burwood, VIC 3125.
Email: rohan.sweeney@deakin.edu.au. Ph: +613 92468543.

Declarations: The authors have no conflicts of interest to declare. Ethics approval was not sort as analyses involved aggregate cross-country data.

Acknowledgements: The data analysed and Stata code used will be made available upon request. Rohan Sweeney received travel support from the University of York via the Alan Williams Fellowship, to facilitate this collaborative research project. The authors would like to thank the anonymous reviewers and Journal Editor who provided invaluable feedback and advice to strengthen the paper. The authors would also like to thank Ryota Nakamura (University of York) and participants at presentations of this work at the London School of Hygiene and Tropical Medicine and the University of York, for their valuable feedback and advice.

The impact of SWAps on health aid displacement of domestic health expenditure.

Keywords: government health expenditure, development assistance, aid additionality, aid fungibility, crowding out, displacement, sector wide approach, aid coordination

Abstract

Recent research suggests that an additional $1 of health aid would displace – or crowd out – nearly the same amount in a recipient governments' own health expenditure. Implementing a Sector Wide Approach (SWAp) may exacerbate crowding out because recipient governments should face fewer constraints when allocating health aid. This paper uses rigorous panel data methods to investigate this hypothesised effect of SWAp. We find that SWAp provides not an exacerbating but a potentially protective effect, reducing displacement of government health expenditure. This suggests some aid dollars are more fungible than others, and the mechanism for aid delivery makes a difference.

1. Introduction

Empirical evidence shows that an additional \$1 of health aid does not lead to an additional \$1 spent on health. Rather Dieleman, Graves, and Hanlon (2013) estimated that over the period 1995 to 2010 an additional \$1 of health aid channelled to government on average reduced that government's own health expenditure from domestic revenues by about \$0.86. This same period saw development agencies, donors, and aid recipients advocate for and achieve dramatic increases in development assistance for health (DAH); ostensibly to assist countries address significant population health problems and health inequalities between (and within) countries (Lu et al., 2010; Ravishankar et al., 2009). The effective treatment of this health aid as almost entirely fungible has been viewed by some as a failure of government to take responsibility for the health of their people, and indeed as a threat to the sustainability of aid supported programmes (Leiderer, 2012; Lu et al., 2010).

The potential for health aid to be less than fully additional arises where the recipient government views aid as a replacement for domestic revenues such as taxation or user fees that would otherwise have been spent on health. The donor allocates aid with the intention of increasing spending on health but a reallocation of domestic funds means that the overall increase in spending is shared between health and other sectors (McGillivray & Morrissey, 2000). This 'reallocation', 'displacement' or 'crowding out' of domestic health expenditure may be a rational response to an increase in health aid when donor and recipient government priorities are misaligned (McGillivray & Morrissey, 2000; Ooms et al., 2010), or an approach to managing fluctuating health aid receipts (Ooms et al., 2010). Nonetheless, donors may seek to influence or constrain recipient allocations if their aid contributions are redirected to priorities they do not share (McGillivray & Morrissey, 2000), or to serve the interests of local 'elites' (Leiderer, 2012).

The available empirical evidence suggests that the potential for health aid to crowd out recipient government health expenditure is very real (Dieleman, Graves & Hanlon, 2013; Dieleman & Hanlon, 2014; Farag, Nandakumar, Wallack, Gaumer, & Hodgkin, 2009; Lu et al., 2010), although Van de Sijpe (2013) argues the extent to which health aid is fungible is exaggerated. However, this same evidence suggests that the extent of crowding out might be modifiable. For example, Lu et al. (2010) report crowding out (of just less than 100%) for DAH directed via recipient governments but a 'flypaper effect' for directly delivered aid. Under the 'flypaper effect', additional government controlled funds 'stick' to increases in DAH such that $1 of directly delivered aid for HIV yields a more than $1 increase in HIV funding (Leiderer, 2012; Pettersson, 2007). A number of previous studies have reported a similar difference in the apparent behavioural response to aid delivered via recipient governments and non-government agents (Dieleman et al., 2013; Farag et al., 2009; Lu et al., 2010).

Theoretical explanations also suggest that the magnitude and direction of any effect may depend upon the mechanism for aid delivery (Leiderer, 2012; Torsvik, 2005). Intuitively, untied grants allow full flexibility for the recipient government to direct grant monies into finance of programs that would have been funded anyway (out of domestic revenues). Tied grants might carry a lower risk of crowding out because donors can designate funds to programs that are not currently being undertaken and that are unlikely to be undertaken by a recipient government left to its own devices. Thus the mode of aid delivery may be particularly important for development partners when considering the opportunity and incentive for displacement.

The most prominent mode of health aid delivery remains project-based (Piva & Dodd, 2009; Ravishankar et al., 2009), the large increases of which post 1990 created very fragmented aid programmes, entailing high transaction costs and reducing aid effectiveness (Buse & Walt, 1997; Hill, 2002; Walt, Pavignani, Gilson & Buse, 1999). Project-based aid has been further criticised for ignoring domestic health priorities and working in parallel rather than in support of government health services, including luring away skilled health care workers (Dieleman et al., 2014; Esser & Bench, 2011; Knack & Rahman, 2007).

A global response to such concerns came in the shape of the Aid Effectiveness Agenda, encapsulated in the Paris Declaration principles for effective aid delivery. Widely endorsed by donors, the Paris Declaration called for greater coordination among donors, who as a group should deliver aid through domestically owned health sector plans, addressing local health priorities, whilst strengthening recipient government capacity to manage aid programmes (Organisation for Economic Cooperation & Development [OECD], 2010). The most prominent formal manifestation of aid effectiveness principles in the health sector has been the Sector Wide Approach (SWAp) (Peters & Chao, 1998; Peters, Paina & Schleimann, 2013; Sundewall & Sahlin-Andersson, 2006).

Whilst the aim of donor coordination via SWAp (or SWAp-like mechanisms) is to improve aid outcomes (eg. in terms of poverty reduction, population health improvements, economic growth), Torsvik (2005) warned that coordinating donors may intensify the crowding out problem. He showed theoretically that, if a unified group of donors is sufficiently concerned with the wellbeing of the poor, the recipient government may be more likely to pursue other priorities with government funds. Conditional (and enforceable) contracts that “re-align” recipient priorities may prevent this.

SWAps have been found to significantly increase levels of health aid delivered as untied sector support (Sweeney & Mortimer, 2016). With this emphasis on untied aid and recipient (rather than donor) control over aid disbursements, it is plausible then that a SWAp may increase crowding out of recipient government health expenditures. However, it is also plausible that SWAps may reduce the likelihood of displacement. Examples exist of where health SWAp agreements have used contracts to cement health expenditure commitments from government and donors alike (Ulikpan, Mirzoev, Jimenez, Malik & Hill, 2014). Further, a decision to implement a health SWAp may represent the commencement of a period of strengthening health sector governance, including the capacity to work with and negotiate with both donors and finance ministries on a more equal footing. Certainly strengthening sector governance is an important aim of SWAps and evidence from case studies suggests that such improvements are achievable (Vaillancourt, 2009).

Hence the net effect of SWAp implementation on the magnitude and direction of crowding out is an empirical question – and one that has hitherto not been assessed. This paper seeks to fill this gap by empirically estimating the magnitude and direction of crowding out or displacement, with and without SWAp. In Section 2, we briefly set out a conceptual framework of displacement. Section 3 outlines our empirical approach, specifying panel data regressions to estimate the potentially moderating effect of SWAps on health aid displacement of government health expenditure using cross-country panel data from 1995-2012. In Sections 4 and 5 we present and discuss the results of the empirical models wherein increases in health aid appear to displace government health expenditure, but at a lower rate in SWAp settings than in non-SWAp settings.

The research has potentially important development policy implications. Many donors wish to curtail the extent of crowding out, concerned that displaced funds are being put to inappropriate, even corrupt use (Kolstad, 2005; Leiderer, 2012).[1] Further, despite persistent underspending on population health, growth in health aid has plateaued; and evidence suggests governments do not replace government expenditure in response to a drop in health aid at the same rate as it is displaced when health aid is increasing (Dieleman & Hanlon, 2014; Leach-Kemon et al., 2012). Even during the period of rapidly increasing health aid post 1990, total health expenditure remained very low in many countries, and was deemed “insufficient to ensure universal access to even a basic set of health services” (World Health Organisation [WHO], 2010, p.31). Thus it is crucial that we gain a better understanding of policies that may help or hinder donors and recipient governments as they attempt to optimise their investments in population health.

2. Background - Theoretical explanation of crowding out

This section provides a brief conceptual framework to illustrate the two opposing theoretical predictions regarding the effect of SWAps on DAH displacement of government health spending. Figure 1 depicts the recipient government's trade-off between expenditures in the health sector versus other sectors. The recipient government commences at point A, with the government's indifference curve (reflecting the relative weight given to public investment in the health sector versus other sectors) tangent to the budget constraint RR. Direct delivery of aid specifically for the health sector (DAH) would shift the budget constraint to the right and full additionality would entail an increase in funding for health by the full amount of the grant (a shift from point A to B). However, in the comparison between point B and other feasible allocations on the new budget constraint, the recipient government can reach a higher indifference curve by allocating the available pool of funds (R+DAH) to reflect their preferences across health sector and other sector activities (I2 at point C), leaving the proportional allocation between health and other sectors exactly as it was before receipt of additional DAH. In this example the allocation of DAH leads to an increase in health sector spending, but not by the full amount of DAH as perhaps donors intended – again reflecting a difference in preferences between donor and recipient government. The extent to which additional DAH is allocated to the health sector will depend on the relative slopes of indifference curves and the budget constraint. This analysis can also be applied at the micro- or meso-level of allocations across disease-areas or specific programmes. For example, DAH tied to HIV programmes may result in government shifting its own spending from HIV to other health sector priorities.

PLEASE INSERT FIGURE 1 ABOUT HERE

Theoretical explanations for the flypaper effect include the presence of non-budgetary constraints that limit the recipient government's ability to disburse funds on particular types of projects (Brooks & Phillips, 2010). Other explanations for the fly-paper effect derive from reductions in the relative prices of goods and services (such as medicines, labour or health services) delivered as aid in-kind but that are also exchanged in domestic markets (McGillivray & Morrissey, 2001). A fall in relative prices may arise if goods and services delivered as aid in-kind are offered at less than market prices, or if aid in-kind is provided in sufficient quantities to force down market prices (McGillivray & Morrissey, 2001). Given a sufficiently large change in relative prices, allocating the available pool of funds in line with recipient preferences may nonetheless result in an increase in health sector spending by more than the full amount of DAH. McGillvray and Morrissey (2001) note that *perceived* changes in relative prices and other types of 'aid illusion' may provoke a similar behavioural response from recipients; providing yet another possible explanation for the fly-paper effect.

3 Empirical Approach

3.1 Main specification

The main specification Equation (1) employs a difference-in-differences approach to identify the moderating effect of SWAp implementation on health aid displacement of government health expenditure. Specifically, we estimate:

$$(1)\quad GHE\text{-}S_{it} = \alpha_i + \beta_1 DAHG_{it} + \beta_2 DAHNG_{it} + \delta_1 SWAp_{it} + \delta_2 SWAp_{it}{*}DAHG_{it} + \delta_3 SWAp_{it}{*}DAHNG_{it} + \beta_3 GGE_{it} + \beta_4 GDP_{it} + \beta_5 HIV_{it} + u_t + \varepsilon_{it}$$

where GHE-S is government health expenditure (sourced from government revenues) in country *i* in year *t.* SWAp is a dummy treatment variable denoting the presence of a health SWAp in country *i* in year *t*. DAHG and DAHNG are development assistance for health (or health aid) channelled to the respective government and non-government sectors of country *i* in year *t*. The main parameter of interest is $\delta_{2,}$ the interaction effect between SWAp and DAHG; capturing the modifying effect of SWAp status on the extent to which DAHG displaces recipient government health expenditure. δ_3 captures the modifying effect of SWAp status on the extent to which DAHNG (DAH that bypasses government) displaces (or increases) government health expenditure.[2] For each interaction term (δ_2 and δ_3), the relevant DAH component has first been demeaned using the sample mean. The resulting interpretation of $\delta_{2,}$ for example, is the impact of increasing DAHG (beyond average levels) on GHE-S when a SWAp is in place.

GGE is general government non-health expenditure, which captures broader changes in government fiscal policies and spending. GDP (per capita) and HIV death rates (per 100,000 population) are included as potential confounding variables, since they influence both levels of government health expenditure as well as need for health expenditure (Dieleman et al., 2013; Dieleman & Hanlon, 2014; Lu et al., 2010; Van de Sijpe, 2013). HIV deaths per 100,000 population are controlled for because some countries with high HIV burden have been committing increased yearly budgets to addressing this health burden (Amico, Aran & Avila, 2010; Lu et al., 2010). Country fixed effects (α_i) and year effects (u_t) capture additional country-specific determinants of health expenditure and global trends. ε is the random error term. All financial variables were included as proportions of GDP, except GDP per capita, which was logged to ease interpretation (as were HIV death rates).

In preliminary robustness analyses, we tested sensitivity to the use of alternative measures of HIV prevalence and included additional controls into the specification. First, infant mortality rate (IMR, per 1,000 live births) was included as an important measure of population health status and thus a measure of health expenditure need. IMR was not included in the main specification because IMR may represent an intermediate outcome in the causal chain between DAH and GHE. Following Lu et al. (2010) and Dieleman et al. (2013) debt relief (which also includes significant missing data) is included as some debt relief initiatives have placed conditions on beneficiaries to increase government expenditures on health and education (Lu et al., 2010). Finally, World Bank governance indicators of perceived government effectiveness and control of corruption are included as they may influence GHE, but also to control for any remaining differences between the SWAp treatment and control groups after sample selection (Kaufmann, Kraay & Mastruzzi, 2010).

3.2 Data sources and limitations

SWAp data was identified via the International Health Partnerships Plus (IHP+) Country Planning Cycle Database (IHP+, 2011) and a search of grey and peer-reviewed literature (see Sweeney, Mortimer & Johnston, 2014a, b). Data for other variables were obtained from IHME, WHO and World Bank databases. The compiled dataset for our main analysis spans 1995-2012. The IHME DAHG and DAHNG data are used because documented attempts are made to differentiate between DAH that is on-budget (i.e. channelled via recipient government), and DAH that is off-budget (i.e. where donors fund endeavours to benefit the recipient country - such as activities of non-government agencies or technical advisors based in the donor country, but these funds are not channelled via the recipient government) (Dieleman, Graves et al. 2013). A number of difficulties in tracking DAH flows and attributing these flows to government or non-government channels have previously been identified in the literature, such that the distinction between on and off-budget aid may not be entirely accurate (Van de Sijpe, 2013a, b). For example, DAHG could include aid that is not associated with, or even known to, the government. Though the extent of measurement error in DAHG and DAHNG data is contested in the literature (e.g. Dieleman et al., 2013; Van de Sijpe, 2013a, b), the possibility of large magnitude errors in identifying off-budget and on-budget aid cannot be excluded (Van de Sijpe, 2013a, b).

For our main analysis, we specify the component of government health expenditure out of domestically sourced revenues (i.e. excluding health aid): GHE-S, as the dependent variable. This is intended to facilitate ready comparison against previous estimates of the fungibility of health aid (Dieleman et al., 2013; Lu et al., 2010). However, GHE-S is a constructed variable, derived by subtracting levels of DAHG from World Health Organisation (WHO) data on government health expenditure as the agent (GHE-A). This is done because GHE-A includes health aid channelled via the government. Following the debate in the literature, this procedure carries the risk of propagating any measurement error in DAHG to the dependent variable. Any such errors in identifying off-budget and on-budget aid have the potential to bias estimated coefficients on DAHG, DAHNG, and their interactions with SWAp. In supplementary analyses, we test sensitivity of our results to measurement error in GHE-S, DAHG and DAHNG (see Section 3.4 and Supplementary Materials for details).

3.3 Sample selection

Countries were included if they had at least 16 of a possible 18 complete annual observations, including a 1995 (baseline) observation. A complete observation was defined as containing data for all of GHE-S, DAHG, DAHNG, SWAp, GDP, and GDP/capita. Zambia was excluded from analysis as their SWAp commenced before the first available GHE-S observation. India and China were excluded from analyses because their immense scale makes a country-wide SWAp very unlikely.

The key underlying assumption in Equation (1) is that both the SWAp implementing and control country groups would have had similar trends in GHE-S in the absence of any SWAp implementation. However, preliminary comparisons of observable baseline (1995) country characteristics reveal that SWAp implementing countries (prior to implementation) were much poorer and with worse health indicators than non-implementing countries (see Supplementary Materials). If trends in GHE-S depend upon initial conditions, then differences at baseline threaten the common trends assumption required for valid estimation using a difference-in-differences approach. To minimise this risk, we select a set of well-matched treatment and control country groups, prior to estimating Equation (1). A linear probability model was used to estimate the propensity for a country to implement a SWAp at any point over the period 1996 to 2012, using 1995 (baseline) levels of government health expenditure (GHE-A), health aid (DAH and number of donors), and proxies for 'need' (GDP/capita, IMR, life expectancy, HIV deaths), as well as population size and geographic regional dummies (see section A of Supplementary Materials for full results). Countries with a propensity score of greater than 0.24 (the lower bound of SWAp implementing countries) were included in the sample, thereby excluding those countries least likely to ever implement a SWAp. Table 1 lists the countries selected for inclusion by this criterion (along with SWAp commencement year). To assess the importance of sample selection decisions, Equation (1) is re-estimated with both tighter and broader (propensity score) inclusion criteria applied, in some cases significantly increasing the size of the control group (see results in Supplementary Materials).

[PLEASE INSERT TABLE 1 ABOUT HERE]

In addition to the rigorous approach taken to identify well-matched treatment and control groups, we investigate likely adherence to the common trends assumption. Given available GHE-A (and consequently GHE-S) observations only commence in 1995, we employ an alternative dataset of public health expenditure that has GHE-A estimates enabling the construction of a panel from 1990 to 1996, prior to any included country's SWAp implementation (International Food Policy Research Institute [IFPRI], 2015). The main limitation of this dataset is that it is only available for a smaller subset of countries. However, it does allow some investigation for any pre-existing differences in GHE-S trends over a seven-year period prior to any included SWAp uptake. Two tests (described and presented in full in Section C of Supplementary Materials) are then undertaken to compare pre-SWAp GHE-S trends between the treatment and control groups.

An endogenous relationship between the SWAp treatment variable and the dependent variable may exist. However, the direction of any potentially resulting bias is not entirely obvious. On the one hand, SWAp has been described as best suited to settings where there is both significant need for aid and also a relatively effective government with sound health sector policies (Brown, Foster, Norton & Nachold, 2001). It is plausible that a country observing higher baseline and ongoing increasing trends in GHE-S may indicate such a setting. In such a case, increasing GHE-S may increase the likelihood of SWAp implementation. Increasing GHE-S may also provide encouragement for donors considering supporting a country's SWAp implementation. Alternatively, steadily increasing GHE-S (driven by increasing expenditure of domestically sourced revenues) will result in DAHG becoming a smaller component of a government's total health expenditure programme (GHE-A). At some point this likely means managing one's health aid programme becomes less burdensome, making SWAp implementation less important and less likely. In any case, great care has been taken to establish well-matched treatment and control groups to minimise selection bias associated with any endogeneity. Table 2 suggests there was no significant difference in baseline levels of GHE-S as a proportion of GDP (there was also no difference in baseline GHE-S/capita) and investigation of pre-SWAp GHE-S trends indicates the treatment and control groups were well matched.

3.4 Investigating robustness

We estimate variations of Equation (1) without SWAp and SWAp interaction variables. This provides a revised estimate of the extent to which DAHG displaces GHE-S, using updated data on DAHG and DAHNG and an expanded dataset (with an extra two years of observations, compared to the most recent published estimates) (Dieleman et al., 2013, Dieleman & Hanlon, 2014). To the extent that these estimates are broadly consistent with results from the existing literature, this analysis may provide something of a test of the external validity of our main specification.

To address concerns regarding measurement error and the inclusion of off-budget aid in DAHG data and / or omission of some off-budget aid from our measure of DAHNG, we re-estimate equation (1) after replacing our dependent variable GHE-S with GHE-A as per Dieleman, Graves et al. (2013). For this specification, the interpretation of β_1, β_2, δ_2 and δ_3 is slightly different. Whereas full additionality (zero crowding out) would imply $\beta_1=0$ when we regress DAHG on GHE-S, we would expect $\beta_1=1$ for regressions on GHE-A, because increases in DAHG will be added dollar-for-dollar to GHE-A in the absence of crowding out (if all DAHG is on budget). Because DAHNG is not reflected in government health expenditure, we would expect $\beta_2=0$ for regressions on both GHE-S and GHE-A in the event of full additionality. While this change in dependent variable prevents the propagation of any measurement error in DAHG data to the left-hand side of the equation that occurs when DAHG is subtracted from GHE-A to obtain GHE-S (Dieleman et al., 2013), it does not address measurement errors in DAHG and DAHNG on the right-hand side (RHS) of the equation. The extent of measurement error in DAHG and DAHNG remains contested in the literature (Dieleman et al., 2013, Van de Sijpe, 2013a, b) and alternative (error-free) measures are not currently available. In such circumstances, errors-in-variables (EIV) models offer a means of correcting for bias due to measurement error in RHS variables. We use (a) the xtewreg command in STATA to implement two-step GMM estimation of the EIV model using higher-order moments or cumulants of residuals to 'correct' for errors in one or more of the RHS variables (Erickson et al., forthcoming; Erickson & Whited, 2002; Erickson & Whited, 2012)[3], and (b) the eivreg command, in which we test sensitivity of results to a range of assumed, predetermined levels of "reliability" of mismeasured variables. Full results are reported in Supplementary Materials (Tables B6-B8).

Equation (1) goes some way in controlling for potential persistence in the dependent variable by including other persistent variables like GDP/capita and GGE as well as including country-specific fixed effects (capturing fixed country characteristics such as climate that predispose to infectious disease epidemics, for example). That said, recent studies of displacement effects of DAHG have included a lagged GHE-S control variable in estimations to control for persistence in GHE-S trends using a System Generalised Method of Moments (SysGMM) approach to control for endogeneity introduced as a consequence of adopting this dynamic specification (Dieleman et al., 2013; Dieleman & Hanlon, 2014). For additional robustness, we estimate dynamic panel models with a lagged GHE-S control using the system GMM approach. Full results are presented in Supplementary Materials.

SWAps have been operationalised differently across settings. As such there remains some uncertainty regarding commencement and classification of some SWAps. The importance of this uncertainty was tested by re-estimation with systematic plausible variation in each SWAp country's assigned year of commencement and exclusion of each SWAp country to assess whether results may be driven by potential misclassification of any one SWAp.

3.5 Investigating heterogeneous SWAp effects

Sweeney et al., (2014a) found that SWAp impacted on funding flows (including DAH levels) differently in the poorest implementing countries and that the SWAp impact on various health and health financing indicators appears to increase as SWAps mature. We follow similar approaches here and investigate (a) if the displacement effect of SWAp in the poorest implementing countries with baseline GDP per capita of one dollar per day or less differs from the broader group of SWAp implementing countries; and (b) if any observed SWAp effect on GHE-S increases in magnitude as country SWAps have time to mature.

Further, much as there exists potential for a given SWAp to become increasingly effective as that particular partnership matures and lessons are learned, there may very well be lessons related to curbing DAH displacement of GHE-S within the SWAp framework. To investigate whether there is evidence to support this idea that "displacement management within SWAp" may have evolved since the early days of SWAp implementation, Equation (1) is re-estimated with the SWAp group broken into subgroups of “early” implementers (countries implementing prior to 2003) and “late” implementers. The year 2003 was selected to separate early from late implementers as it represents an obvious break in the data as well as the timing of the First High Level International Forum on Aid Effectiveness in Rome – an opportunity for “early” implementers to share their experiences with SWAp. Both “early” and “late” SWAp effects were modelled using the full set of control countries.

Finally, the corrupt use of displaced funds has been reported as a significant donor concern in this literature (Kolstad 2005; Leiderer, 2012). We explore the potential that SWAp may have heterogeneous effects, depending upon potential for corrupt use of funds, by again splitting the sample into two subgroups, now based upon baseline scores of the World Bank’s governance indicator for “control of corruption” (Kaufmann et al., 2010) where baseline scores were imputed using the first available observation (1996 for most countries). A threshold was set at the index score of -0.76 based upon the presence of an obvious break in the data, quite clearly splitting the control country group in two. Baseline scores less than -0.76 would represent relatively poorer control of corruption. An alternative threshold was applied in supplementary analyses to assess the sensitivity of results to this decision (see Supplementary Materials, Table D2).

4. Results

Table 2 reports baseline characteristics of the two groups. The groups appear well matched, though the control group had significantly higher non-health general government expenditure (GGE/GDP). The presence of such potential differences makes it important to ensure they are controlled for in the main model.

[PLEASE INSERT TABLE 2 ABOUT HERE]

To first test the common trends assumption (discussed in Section 3.3), we analysed an alternative dataset over the period 1990 to 1996 (WHO GHE-S data is not available for this period), to examine if there was any evidence to suggest there were pre-existing differences in GHE-S trends between the treatment (SWAp countries) and control groups prior to any included SWAp implementation. Using public health expenditure data from the International Food Policy Research Institute (IFPRI, 2015), we estimated equation (1) after replacing SWAp variables with (i) the interaction between SWAp and a linear time trend to distinguish the time trend in SWAp countries from the time trend in non-SWAp countries (modelled as year dummies), and (ii) a dummy variable to simulate introduction of a fake SWAp to all SWAp group countries in 1993. Both tests attempt to identify any pre-existing differences in GHE-S trends between the two groups, which would undermine the common trends assumption underpinning the estimation approach and would raise an issue of selection bias. Neither test detected a significant difference between the SWAp and non-SWAp countries, providing support for the common trends assumption (see section C of Supplementary Materials for full results).

Turning now to the main model (column 1, Table 3), we found that a one dollar increase in DAHG was associated with a decrease in GHE-S by about $0.78; consistent with a significant displacement effect. DAHNG had no significant impact on GHE-S; consistent with full additionality. Of primary interest though is that the delivery of an additional one dollar of DAHG (beyond average levels) to a SWAP implementing country rather than a non-implementing country was associated with an increase in GHE-S of about $0.52. This suggests increasing DAHG (beyond average levels of DAHG) by one dollar to a SWAp setting displaces GHE-S by about $0.26 (obtained by summing main and interaction effects: -$0.78 + $0.52c = -$0.26). Further, implementing a SWAp was itself associated with a statistically significant increase in GHE-S, though the magnitude of effect was very small.

Unsurprisingly, increases in general government (non-health) expenditure were also associated with significant increases in GHE-S, suggesting the health sector is included in general fiscal expansions. Interestingly, an increase in control of corruption was also associated with an increase in GHE-S. Both GHE-S and control of corruption may represent some indirect measure of government fiduciary responsibility, and this positive relationship may be intuitively plausible.

[PLEASE INSERT TABLE 3 ABOUT HERE]

The main effect was robust to a suite of additional sensitivity tests presented more fully in Supplementary Materials (section B). Firstly, the significant protective effect of SWAp on DAHG displacement of GHE-S remained present when estimated in a dynamic panel structure using a system GMM approach, though the magnitude of the effect was smaller. Second, results from the EIV models (see Supplementary Materials, tables B6-B8 and accompanying text) confirm that headline results from our main model remain plausible under a range of assumptions regarding measurement error, but that further work is required before drawing strong conclusions. Other analyses included re-estimation using alternative sample selection criteria, plausible variation in SWAp commencement times and systematic exclusion of individual SWAps (thereby excluding chance that results were driven by an individual misclassified SWAp country). The SWAp protective effect remained robust to these analyses.

When the sample was restricted to the subgroup of "poorest" countries (with baseline GDP/capita of no more than $1/day), the protective effect of SWAp remained largely unchanged (see column 1, Table 4). However, it appears there is a slightly larger non-SWAp displacement effect of DAHG on GHE-S in this subsample, perhaps increasing the relative importance of SWAp as a strategy for reducing the effects of displacement. These results were robust to the inclusion of additional variables and alternative sample selection criteria (see Table D1 in Supplementary Materials). It is also worth noting that GDP/capita has a significant and negative association with GHE-S in this subgroup of countries. Further research may indicate whether this might be related to GHE-S declining in response to increased out-of-pocket expenditure on health.

[PLEASE INSERT TABEL 4 ABOUT HERE]

In other investigations of heterogeneity presented in Table 4, it does not appear as though the displacement effect of SWAp on DAH allocated to government has changed significantly between the early period of SWAp implementation and the later period. However, channelling DAH via non-government actors now results in a significant fly-paper effect when a SWAp is in place (columns 2 & 3, Table 4). It also appears that the effect of SWAp on DAH displacement may vary, depending upon control of corruption. Specifically, the protective effect of SWAp on the displacement of GHE-S when DAH is channelled to government appears higher in the subgroup of countries with relatively better control of corruption at baseline (columns 4 & 5, table 4). Interestingly, in the subgroup of countries with poorer control of corruption at baseline, there is evidence to suggest channelling DAH via non-government actors may also reduce the extent of displacement of GHE-S when those funds are allocated to a SWAp setting (column 4, Table 4). More research is required to tease out and better understand these preliminary findings.

Table 5 presents the estimation results for the evolving SWAp effect. Whilst the direction of SWAp on displacement was difficult to predict a priori, there was an expectation that as SWAps mature, the magnitude of influence might increase. This prior is confirmed by the results in Table 5. As a SWAp matures beyond 3 years, its protective effect on GHE-S increases – eventually peaking in the most mature SWAps where a $1 increase in DAHG to a SWAp country increases GHE-S by $0.64 (for a net displacement effect of $0.17 as compared to displacement of $0.81 per $1 in non-SWAp settings). The presence of a significant difference in the pre-SWAp period is of interest. One cannot entirely exclude the possibility that this indicates there may have been some pre-existing difference in GHE-S trends between the SWAp and control group, though earlier analyses of the common trends assumption did not hint at such a difference. Alternatively, this could reflect a preparatory effect, in that the decision to move towards SWAp implementation may itself have resulted in the recipient country government taking its commitments to the health system (as indicated by expenditure) more seriously.

[PLEASE INSERT TABLE 5 ABOUT HERE]

6. Discussion

In light of the ex ante ambiguous effect of a health SWAp on the effect of health aid on government health expenditure, we have empirically assessed what the net effect might be. While theoretically it seemed plausible that the increased control over how health aid can be allocated – afforded by SWAp – might increase the extent of displacement, our results reject this prediction. By contrast, increasing health aid where a SWAp is in place appears to reduce the magnitude of any displacement. The main model predicted that allocating an additional dollar of health aid to a country where a SWAp is in place may reduce the magnitude of displacement from about $0.78 (where there is no SWAp) to about $0.26. Or to put it another way, $0.52 more cents of domestic government revenues is spent in the health sector when health aid is channelled to settings where there is a SWAp in place. So whilst a SWAp cannot facilitate the end of health aid fungibility, it does appear to reduce it significantly. Moreover, implementing a SWAp was itself associated with a statistically significant increase in GHE-S, though the magnitude of effect was very small. Whilst there is some dispute in the literature as to the extent to which the available DAH data captures off-budget aid, even if displacement is overestimated as argued by Van De Sijpe, (2011a, b), the analysis still suggests SWAp is beneficial for increasing health spending.

This analysis also provides an updated estimate of the displacement effect of DAHG on GHE-S over an extended period of analysis and using updated DAHG data from the IHME. Reassuringly, the displacement estimates of DAHG on GHE-S are very consistent with those by Dieleman et al. (2013), despite using an alternative estimation approach as well as a different specification. This provides some confidence as to the external validity of our model.

Previous theoretical and empirical works have suggested that contracts or agreements might be used to produce a flypaper effect, where government funds stick to aid dollars. This appears to be one possible explanation for the protective effect of SWAp in reducing the extent of displacement. For example, the Kyrgyz government was expected to increase state expenditure by 0.6% annually as a condition of continued donor participation in the Kyrgysztan SWAp (Ulikpan et al., 2014).

A focus of SWAp evaluations understandably lay in examining changes in relationship dynamics between a government (usually health ministry) and donors, rather than the health ministry with other government ministries. Nonetheless, case-studies provide a good deal of indirect evidence that increased capacity of health sector officials to negotiate and retain budgetary share under SWAp may have contributed to the observed finding. For example, case-study evidence suggests that recipient government health sector leadership has become increasingly influential over SWAp policy setting and funding decisions as the evaluated SWAps have matured (Foster et. al., 2000; Jefferys & Walford, 2003 ; Vaillancourt, 2009; Walford, 2007). Specifically, Vaillancourt (2009) evaluated health SWAps in Bangladesh, Ghana, Kyrgyz Republic, Malawi, Nepal, Tanzania and concluded that all but

Malawi had made substantial progress on establishing country-led partnerships and increasing health sector leadership. Such conclusions are consistent with the findings of this paper that the extent to which SWAps reduce displacement, appears to increase as SWAps mature – capacity building takes time. Peterset al. (2013) however, concluded that the extent to which SWAps have increased health sector leadership has been mixed.

Case-studies also provide limited direct evidence for an increased capacity of health sector officials to negotiate and retain budgetary share under SWAp. Walford (2007) describes how SWAPs have been used to help facilitate better links between the Ministries of Health and Finance and help Ministries of Health develop their case for funding. When Zambia's major donors reportedly shifted aid delivery from DAH to general budget support (untied to any sector) as a "reward" for strong SWAp and DAH management, there was concern that overall health sector funding would fall (Sundewall et al., 2010; Walford, 2007). However, with the United Kingdom's (DFID) support, the Ministry of Health negotiated with the Ministry of Finance to secure a larger budget to replace the reduction in DFID's DAH (Sundewall et al., 2010; Walford, 2007). It should be kept in mind that the majority of evidence of health SWAp impacts is drawn from case studies where the lack of a counterfactual makes it difficult to control for potentially important confounding effects.

In an important contribution to the discussion on aid fungibility, McGillivray and Morrissey (2000) raised the very fundamental question of whether fungibility even mattered. They suggested that what was ultimately more important was maximising welfare gains from government and aid expenditure taken together and it may be less important to worry about which funds were allocated where. Whilst we are sympathetic to this view, not all donors attach equal weight to welfare gains in health and non-health sectors. Moreover, allocation of displaced funds to other sectors may be driven by political considerations rather than by efforts to maximise an overarching social welfare function (Hauck & Smith, 2015).

From the perspectives of various stakeholders, there is a lot to suggest that displacement does indeed 'matter' considerably. Certainly bilateral donors have previously noted the perceived corrupt use of displaced DAH (Leiderer 2012, Kolstad 2005). Whilst displacement does not imply corrupt use (funds may be put to an alternative, potentially welfare increasing use) (Morrissey 2015), the WHO and other development agencies remain concerned where displacement occurs in poorer aid recipient countries, many of whom are unable to provide even a basic set of health services for their populations (WHO 2010). For the same reason, displacement (or at least government expenditure on health) appears to matter greatly to the recipient nation's people, many of whom face significant financial hardship, financing health care out of pocket (WHO, 2010). Further, whilst we cannot know if spending that marginal DAHG dollar (effectively a marginal public revenue dollar) on health is welfare maximising, there is at least evidence to suggest that a marginal DAH dollar can be welfare increasing, yielding improvements in population health as measured by for example: immunisation rates (Dietrich, 2011; Feeny & Ouattara, 2013), disease-prevalence (Yogo & Mallaye, 2015), and infant or child mortality (Kizhakethalackal et al., 2013; Mishra & Newhouse, 2009; Mukherjee & Kizhakethalackal 2013, Yogo & Mallaye 2015).

Additional findings from this analysis suggest that displacement of government expenditure may be greater in poorest countries; perhaps increasing the value of supporting SWAps in such settings; and that the presence of a SWAp may reduce the extent of GHE-S displacement of (a) a marginal dollar of DAH to government when allocated to a setting with relatively stronger (baseline) control of corruption and (b) a marginal dollar of DAH channelled via non-government actors when allocated to a setting with relatively weaker (baseline) control of corruption. Further investigations are required to better understand this relationship, but these results support the interpretation that the size of the moderating effect of SWAp on health aid displacement could well differ across settings. Hence, donors may need to consider this in deciding the nature of their engagement with SWAp in different contexts.

We believe our findings are particularly timely now, at a moment that sees health SWAps at something of a crossroads and seemingly, donor enthusiasm for the Paris Principles, waning (Peters et al., 2013; Shorten et al., 2012). Donor decisions to refocus on delivering project–based aid with the hope that maintaining control over DAH allocations will (a) reduce recipient opportunity to displace funds (as some might perceive would be present under a SWAp); and (b) enable identification of funding opportunities that limit displacement of GHE-S or even create a flypaper effect, and may seem like an appropriate policy response. However, our findings suggest the opposite may be the case. Empirically, it appears that re-engaging with the principles for effective aid delivery via a SWAp (or a SWAp-like mechanism), may be a means to help donors reduce the extent to which their health aid contributions are treated as a replacement for domestic government health expenditure commitments.

In interpreting our findings, it is important to acknowledge the limitations of the work. Van de Sijpe (2013a, b) demonstrated convincingly the potential for upward bias in estimates of the fungibility of health aid due to inclusion of off-budget aid in DAHG data. While the extent of measurement error in DAHG and DAHNG data is contested in the literature (Dieleman et al., 2013; Van de Sijpe, 2013a, b), tracking DAH flows is a difficult task and there remains a risk that existing measures of DAHG include non-negligible levels of off-budget aid. Moreover, where SWAp improves traceability of DAHG or brings previously off-budget aid on to recipient government budgets, there is potential for a between-group difference in measurement error to arise; resulting in upward bias in estimates of the moderating effect of SWAp on crowding out. While bringing a larger share of aid on budget would be a desirable and important effect of SWAp, this is not the mechanism via which SWAp is hypothesised to influence GHE-S in the present study.

As such, we have made significant efforts to distinguish between the moderating effect of SWAp on crowding out and reductions in the extent of measurement error under SWAp. Specifically, we have conducted robustness tests using errors-in-variables (EIV) models to correct for bias due to measurement error in DAHG, DAHNG, DAHG*SWAp and DAHNG*SWAp. While our headline results from standard FE and system GMM models remain plausible under a range of assumptions regarding measurement error, such corrections constitute second-best solutions. Our findings should be treated as provisional until replicated using data for which consensus has been reached that “an accurate and transparent delineation of on- and off-budget aid” (Van de Sijpe, 2013b) has been achieved. That said, information provided by Dieleman et al., (2013) describing the IHME approach to tracking on- and off-budget aid, indicates improvements have been made in the reliability of data to that used in the earlier papers, which motivated the important concerns raised by Van de Sijpe (2013a).

With regards to general limitations of the approach employed in the present study, Ooms et al. (2010) previously argued that understanding why health aid displaces government health expenditure is as important as knowing the average size of displacement. This study could be equally criticised as we only estimate average effects, missing out on any nuance of effect as SWAps operationalised differently will likely have achieved different effects. Whilst we explore some heterogeneity and context effects in our extensive range of robustness tests, further research is required to tease out lessons about which core SWAp elements are most important to realise its protective effects. Currently such detailed data on SWAp implementation by country does not exist. In related work, we are currently exploring different quantitative approaches to estimate individual country impacts of SWAp (eg. employing difference-in-differences analyses for individual SWAp implementers; and the panel data approach to policy evaluation set out in Hsiao, Ching & Wan (2012)). It may be possible to identify particularly “strong” SWAps using this approach, motivating use of case study methods to understand the likely factors that have led to “strong” performance.

Significant steps were taken to remove sources of selection bias and we believe the SWAp implementing and control groups are well matched. However, it is possible that some selection bias persists. As such there remains a risk that between group differences in additional unobserved factors may be influencing the extent of change in GHE-S attributed to SWAp. Baseline comparisons and tests of common trends in GHE-S prior to the period of implementation provide some comfort that the groups were well matched; yet data limitations over the period prior to significant SWAp uptake (1990 to 1996) do introduce some uncertainty.

7. Conclusion

Health aid remains a vital source of funds for a significant sub-group of recipient countries (Lu et al., 2010; WHO, 2010). The countries included in our analysis includes the poorest of the poor, with health sectors nowhere near sourcing sufficient domestic funds for a minimum package of essential services. For these countries, high levels of fungibility undermine efforts to bridge the funding gap. This analysis provides support that allocating health aid to governments where a SWAp is in place may help by reducing the extent of displaced government health expenditure. This may be as a result of contracts within SWAp securing recipient government contributions. It may also be as a result of strengthening health sector governance, increasing capacity to secure government funds for the sector. Irrespective of the specific mechanism, this potential benefit of SWAp should be given weight by donors and health ministries considering – or already in the process of – managing their health aid programmes within a health SWAp.

Endnotes:

[1] Morrissey (2015) notes that some donors hold a "skeptical view" of general budget support (GBS); primarily due to concerns that GBS may carry a heightened risk of corruption. However, Morrissey (2015) argues that corruption under GBS should be

less of a concern given more recent studies show improved, more transparent recipient financial management systems and that GBS further supports such improvements (p.103).

[2] Interpretation of β_1, β_2, δ_2 and δ_3 as described above is predicated on valid and reliable measurement of *DAHG* and *DAHNG.* A number of difficulties in tracking DAH flows and attributing these flows to government or non-government channels have previously been identified in the literature (Van de Sijpe (2013a, b). The potential for such difficulties to bias estimates of model parameters is noted and discussed in Section 3.2 and 3.4 of the main manuscript and in Supplementary Materials.

[3] Erickson & Whited (2012) run Monte Carlo experiments to evaluate the finite sample properties of this and other approaches to addressing measurement error. Importantly, the use of higher-order moments/cumulants in two-step GMM is robust to serial correlation in the measurement errors. In contrast, using lags of mismeasured regressors as instruments in IV or dynamic panel models "can produce the same biased results as does OLS if the measurement error is serially correlated" (Erickson & Whited, 2012 p1287). Previous attempts to address measurement error in aid fungibility models using lags in dynamic panel models (Dieleman et al., 2013) were criticised by Van de Sijpe (2013b) for their potential sensitivity to serial correlation in measurement errors.

REFERENCES

Amico, P., Aran, C. & Avila, C. (2010). HIV spending as a share of total health expenditure: An analysis of regional variation in a multi-country study. *PLoS ONE, 5,* e12997.

Brooks, L. & Phillips, J. H. (2010). An institutional explanation for the stickiness of federal grants. *Journal of Law, Economics, and Organization, 26,* 243-264.

Brown, A., Foster, M., Norton, A., & Nachold, F. (2001). The status of sector wide approaches. Overseas Development Institute.

Buse, K. (1999). Keeping a tight grip on the reins: donor control over aid coordination and management in Bangladesh. *Health Policy and Planning,14,* 219-228.

Buse, K. & Walt, G. (1997). An unruly melange? Coordinating external resources to the health sector: a review. *Social Science and Medicine, 45,* 449-463.

Chansa, C., Sundewall, J., McIntyre, D., Tomson, G., & Forsberg, B. C. (2008). Exploring SWAp's contribution to the efficient allocation and use of resources in the health sector in Zambia. *Health Policy and Planning, 23,* 244-251.

Dieleman, J. L., Graves, C. M., & Hanlon, M. (2013). The fungibility of health aid: Reconsidering the reconsidered. *The Journal of Development Studies, 49,* 1755-1762.

Dieleman, J. L., Graves, C. M., Templin, T., Johnson, E., Baral, R., Leach-Kemon, K., …Murray, C. J. (2014). Global health development assistance remained steady in 2013 but did not align with recipients' disease burden. *Health Affairs (Millwood), 33,* 878-886.

Dieleman, J. L., & Hanlon, M. (2014). Measuring the displacement and replacement of government health expenditure. *Health Economics, 23,* 129-140.

Dietrich, S. (2011). The politics of public health aid: Why corrupt governments have incentives to implement aid effectively. *World Development, 39,* 55-63.

Erickson, T., & Whited, T. (2002). Two-step GMM estimation of the errors-in-variables model using high-order moments. *Econometric Journal, 18,* 776-99.

Erickson, T., & Whited, T. (2012). Treating measurement error in Tobin's q. *The Review of Financial Studies, 25,* 1286-1329.

Erickson, T., Parham, R., & Whited, T. (forthcoming). xtewreg: Estimating the errors-in-variables model using high-order cumulants and moments of arbitrary degree. *Stata Journal,* forthcoming.

Esser, D. E., & Keating Bench, K. (2011). Does global health funding respond to recipients' needs? Comparing public and private donors' allocations in 2005–2007. *World Development, 39,* 1271-1280.

Farag, M., Nandakumar, A. K., Wallack, S. S., Gaumer, G., & Hodgkin, D.(2009). Does funding from donors displace government spending for health in developing countries? *Health Affairs, 28,* 1045-1055.

Feeny, S., & Ouattara, B. (2013). The effects of health aid on child health promotion in developing countries: cross-country evidence. *Applied Economics, 45,* 911-919.

Foster, M., Brown, A., Conway, T. (2000). Sector-wide approaches for health development: A review of experience. World Health Organization, Global Programme on Evidence for Health Policy. http://whqlibdoc.who.int/hq/2000/WHO_GPE_00.1.pdf.

Hauck, K., & Smith, P.C. (2015). The politics of priority setting in health: A political economy perspective. *Centre for Global Development*. Working paper 414, September 2015. https://www.ciaonet.org/attachments/27959/uploads?1442194766. Date accessed, May 12th, 2016.

Hill, P. S. (2002). The rhetoric of sector-wide approaches for health development. *Social Science and Medicine, 54,* 1725-1737.

Hsiao, C., Ching, H. S., & Wan, S. K. (2012). A panel data approach for program evaluation: measuring the benefits of political and economic integration of Hong Kong with mainland China. *Journal of Applied Econometrics, 27,* 705-740.

IFPRI. (2015). Statistics on public expenditures and economic development (SPEED). International Food Policy Research Institute. Washington, D.C.

IHP+ (2011). Country planning cycle database: a WHO resource, International Health Partnerships Plus. www.nationalplanningcycles.org.

Jefferys, E. & Walford, V. (2003). Mapping of sector wide approaches; a report for the Swedish International Development Cooperation Agency. Institute for Health Sector Development. London. http://www.swisstph.ch/fileadmin/user_upload/Pdfs/swap/swap319.pdf.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010). The worldwide governance indicators: methodology and analytical issues. *Policy Research Working Paper* No. 5430. World Bank: Development Research Group. Washington DC.

Kizhakethalackal, E. T., Mukherjee, D., & Alvi, E. (2013). Quantile regression analysis of health-aid and infant mortality: A note. *Applied Economics Letters, 20,* 1197-1201.

Kolstad, I. (2005). Direct budget support and corruption. U4 Anti-Corruption Resource Centre, Michelsen Institute, Bergen, U4 Issue 1:2005.

Knack, S., & Rahman, A. (2007). Donor fragmentation and bureaucratic quality in aid recipients. *Journal of Development Economics, 83,* 176-197.

Leach-Kemon, K., Chou, D. P., Schneider, M. T., Tardif, A., Dieleman, J. L., Brooks, B. P. C., …Murray, C. J. L. (2012). The global financial crisis has led to a slowdown in growth of funding to improve health in many developing countries. *Health Affairs, 31,* 228-235.

Leiderer, S. (2012). Fungibility and the choice of aid modalities: the red herring revisited. Helsinki, United Nations University World Institute for Development Economics Research.

Lu, C., Schneider, M. T., Gubbins, P., Leach-Kemon, K., Jamison, D., & Murray, C. J. L. (2010). Public financing of health in developing countries: a cross-national systematic analysis. *The Lancet, 375,* 1375-1387.

McGillivray, M., & Morrissey, O. (2000). Aid fungibility in Assessing Aid: red herring or true concern? *Journal of International Development, 12,* 413-428.

McGillivray, M., & Morrissey, O. (2001). Aid illusion and public sector behaviour. *Journal of Development Studies, 37,* 118-136.

Mishra, P., & Newhouse, D. (2009). Does health aid matter? *Journal of Health Economics, 28,* 855-872.

Morrissey, O. (2015). Aid and government fiscal behavior: Assessing recent evidence. *World Development, 69,* 98-105.

Mukherjee, D., & Kizhakethalackal, E. T. (2013). Empirics of health-aid, education and infant mortality: a semiparametric study. *Applied Economics*, *45,* 3137-3150.

OECD (2010). Paris Declaration and Accra Agenda for Action. OECD Development Co-operation Directorate.

Ooms, G., Decoster, K., Miti, K., Rens, S., Van Leemput, L., Vermeiren, P., & Van Damme, W. (2010). Crowding out: are relations between international health aid and government health funding too complex to be captured in averages only? *Lancet, 375,* 1403-1405.

Peters, D., & Chao, S. (1998). The sector-wide approach in health: What is it? Where is it leading? *The International Journal of Health Planning and Management, 13,* 177-190.

Peters, D. H., Paina, L., & Schleimann, F. (2013). Sector-wide approaches (SWAps) in health: what have we learned? *Health Policy and Planning, 28,* 884-890.

Pettersson, J. (2007). Foreign sectoral aid fungibility, growth and poverty reduction. *Journal of International Development, 19,* 1074-1098.

Piva, P., & Dodd, R. (2009). Where did all the aid go? An in-depth analysis of increased health aid flows over the past 10 years. *Bulletin of the World Health Organization, 87,* 930-939.

Ravishankar, N., Gubbins, P., Cooley, R. J., Leach-Kemon, K., Michaud, C. M., & Jamison, D. T. (2009). Financing of global health: tracking development assistance for health from 1990 to 2007. *Lancet, 373,* 2113-2124.

Shorten, T., Taylor, M., Spicer, N., Mounier-Jack, S., & McCoy, D. (2012). The International Health Partnership Plus: Rhetoric or real change? Results of a self-reported survey in the context of the 4th high level forum on aid effectiveness in Busan. *Globalization and Health*, *8,* 13.

Sundewall, J., Jonsson, K., Cheelo, C., & Tomson, G. (2010). Stakeholder perceptions of aid coordination implementation in the Zambian health sector. *Health Policy, 95,* 122-128.

Sundewall, J., & Sahlin-Andersson, K. (2006). Translations of health sector SWAps--a comparative study of health sector development cooperation in Uganda, Zambia and Bangladesh. *Health Policy, 76,* 277-287.

Sweeney, R., & Mortimer, D. (2016). Has the SWAp influenced aid flows in the health sector? *Health Economics, 25,* 559-77.

Sweeney, R., Mortimer, D., & Johnston, D. W. (2014a). Do sector wide approaches for health aid delivery lead to 'donor-flight'? A comparison of 46 low-income countries. *Social Science and Medicine, 105,* 38-46.

Sweeney, R., Mortimer, D., & Johnston, D. W. (2014b). Further investigations of the donor-flight response. *Social Science and Medicine, 113,* 179-182.

Torsvik, G. (2005). Foreign economic aid; should donors cooperate? *Journal of Development Economics, 77,* 503-515.

Ulikpan, A., Mirzoev, T., Jimenez, E., Malik, A., & Hill, P. S. (2014). Central Asian Post-Soviet health systems in transition: has different aid engagement produced different outcomes? *Global Health Action, 7,* 24978.

Vaillancourt, D. (2009). Do health sector-wide approaches achieve results? Emerging evidence and lessons from six countries. IEG Working Paper 2009/4. World Bank. Washington.

Van de Sijpe, N. (2013a). The fungibility of health aid reconsidered. *Journal of Development Studies, 49,* 1746-1754.

Van de Sijpe, N. (2013b). The fungibility of health aid reconsidered: A rejoinder. *Journal of Development Studies, 49,* 1763-1764.

Walford, V. (2003). Defining and evaluating SWAps: A paper for the Inter-Agency Group on SWAps and development cooperation. Institute for Health Sector Development. London. http://www.hlsp.org/Home/Resources/DefiningevaluatingSWAps.aspx.

Walt, G., Pavignani, E., Gilson, L., & Buse, K. (1999). Managing external resources in the health sector: are there lessons for SWAps (sector-wide approaches)? *Health Policy and Planning, 14,* 273-284.

WHO. (2010). "World health report: Health financing the path to universal coverage. WHO Press, Geneva.

Yogo, U. T., & Mallaye, D. (2015). Health aid and health improvement in sub-Saharan Africa: Accounting for the heterogeneity between stable states and post-conflict states. *Journal of International Development, 27,* 1178-1196.

Table 1. Countries included in main sample (with year of SWAp commencement)

SWAp implementing countries (n=27)		Non-implementing control countries (n=25)	
Bangladesh (1997)	Mozambique (1997)	Angola	Honduras
Benin (2010)	Nepal (2004)	Armenia	Indonesia
Burkina Faso (2001)	Nicaragua (2005)	Cape Verde	Kenya
Burundi (2008)	Niger (2006)	Central African Republic	Kiribati
Cambodia (2000)	Papua New Guinea (2008)	Chad	Laos
Ethiopia (1997)	Rwanda (2007)	Comoros	Nigeria
Ghana (1997)	Senegal (1997)	Congo	Pakistan
Kyrgyzstan (2005)	Solomon Is. (2009)	Congo, DRC	Philippines
Lesotho (2005)	Sudan (2006)	Eritrea	Sierra Leone
Madagascar (2007)	Tanzania (1997)	Fiji	Sri Lanka
Malawi (2004)	Uganda (2000)	Guinea	The Gambia
Mali (1998)	Uzbekistan (2009)	Guinea-Bissau	Togo
Mauritania (2010)	Vietnam (2009)	Haiti	
Mongolia (2009)			

Table 2. Comparison of baseline (1995) characteristics.

	SWAp Countries					Control Countries					
Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**	T-test p-value
GDP/capita ($)	27	411.71	301.49	134.40	1445.66	25	590.96	502.36	134.33	2540.46	0.13
Population (millions)	27	18.89	25.95	0.36	119.87	25	26.26	48.47	0.08	194.11	0.50
GHES/GDP (%)	27	1.33	1.42	-2.71	3.69	25	1.17	2.85	-8.04	8.68	0.81
DAHG/GDP (%)	27	0.83	0.95	0.00	3.74	25	1.21	2.40	0.00	10.82	0.46
DAHNG/GDP (%)	27	0.29	0.30	0.00	1.21	25	0.19	0.21	0.00	0.79	0.19
GGE/GDP (%)	27	20.47	8.51	5.51	52.25	25	28.40	18.81	6.62	77.21	0.06
IMR (per 1,000 live births)	27	84.39	29.57	29.10	143.40	25	80.90	37.66	17.20	153.40	0.71
Life expectancy	27	54.99	8.68	31.24	72.14	25	56.06	9.43	35.82	69.42	0.67
HIV deaths (per 100,000)	27	76.12	119.66	0.00	424.50	25	62.10	106.41	0.00	456.00	0.66

Table 3. Impact of SWAp on aid displacement

	(1) GHE-S	(2) GHE-S	(3) GHE-S	(4) GHE-S	(5) GHE-S
$DAHG_{it}$	-0.779***	-0.779***	-0.777***	-0.762***	-0.762***
	(0.061)	(0.060)	(0.074)	(0.080)	(0.116)
$SWAp_{it}$	0.005***	0.005***	0.005***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$SWAp*DAHG_{it}$	0.518***	0.519***	0.514***	0.441***	0.440***
	(0.123)	(0.122)	(0.124)	(0.124)	(0.127)
DAHNG	0.124	0.124	0.098	0.098	0.075
	(0.103)	(0.103)	(0.118)	(0.096)	(0.111)
SWAP*DAHNG	-0.084	-0.086	-0.063	-0.063	-0.050
	(0.198)	(0.203)	(0.202)	(0.192)	(0.200)
Log (GDP/capita) $_{it}$	-0.003	-0.003	-0.003	-0.002	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
GGE_{it}	0.029***	0.029***	0.029***	0.025**	0.025**
	(0.010)	(0.010)	(0.010)	(0.011)	(0.012)
Log (HIV deaths)$_{it}$[a]	0.000	0.000	-0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log (IMR)$_{it}$		-0.000			-0.001
		(0.005)			(0.005)
Debt relief $_{it}$			0.002		0.001
			(0.004)		(0.003)
Government effectiveness $_{it}$				0.000	0.000
				(0.002)	(0.002)
Corruption Control $_{it}$				0.004**	0.004*
				(0.002)	(0.002)
N	936	936	918	878	863
SWAp countries	27	27	27	27	27
Control countries	25	25	24	25	24

Standard errors in parentheses; * p < 0.10, ** p < 0.05, *** p < 0.01; year trends included, results suppressed for brevity.
[a] The main model (column 1) was also robust to re-estimation using HIV prevalence data from the IHME and World Bank – see Table B9, Supplementary Materials. Both of these datasets contain missing data, thus HIV deaths has been used in the main model.

Table 4. Heterogeneous effects of SWAp on DAH displacement of GHE

	(1) Poorest subgroup[a]	(2) Early implementers[b]	(3) Late implementers[b]	(4) Poorer control of corruption[c]	(5) Stronger control of corruption[c]
	GHE-S	GHE-S	GHE-S	GHE-S	GHE-S
$DAHG_{it}$	-0.860***	-0.856***	-0.781***	-0.798***	-0.762***
	(0.063)	(0.036)	(0.063)	(0.064)	(0.105)
$SWAp_{it}$	0.006***	0.002	0.006***	0.005**	0.004***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
$SWAp*DAHG_{it}$	0.543***	0.346**	0.399***	0.253	0.544***
	(0.093)	(0.150)	(0.146)	(0.210)	(0.140)
DAHNG	0.127	0.037	0.155	0.135	0.141
	(0.155)	(0.117)	(0.107)	(0.205)	(0.086)
SWAP*DAHNG	-0.143	-0.139	0.296*	0.408*	-0.308
	(0.232)	(0.221)	(0.153)	(0.213)	(0.215)
Log (GDP/capita) $_{it}$	-0.007**	-0.004*	-0.003	-0.002	-0.003
	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)
GGE_{it}	0.028**	0.029***	0.030***	0.032*	0.024***
	(0.012)	(0.010)	(0.010)	(0.016)	(0.006)
Log (HIV deaths)$_{it}$	-0.001	-0.001	0.000	-0.001	0.001
	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
N	504	648	738	476	476
SWAp countries	17	11	16	17	17
Control countries	11	25	25	11	11

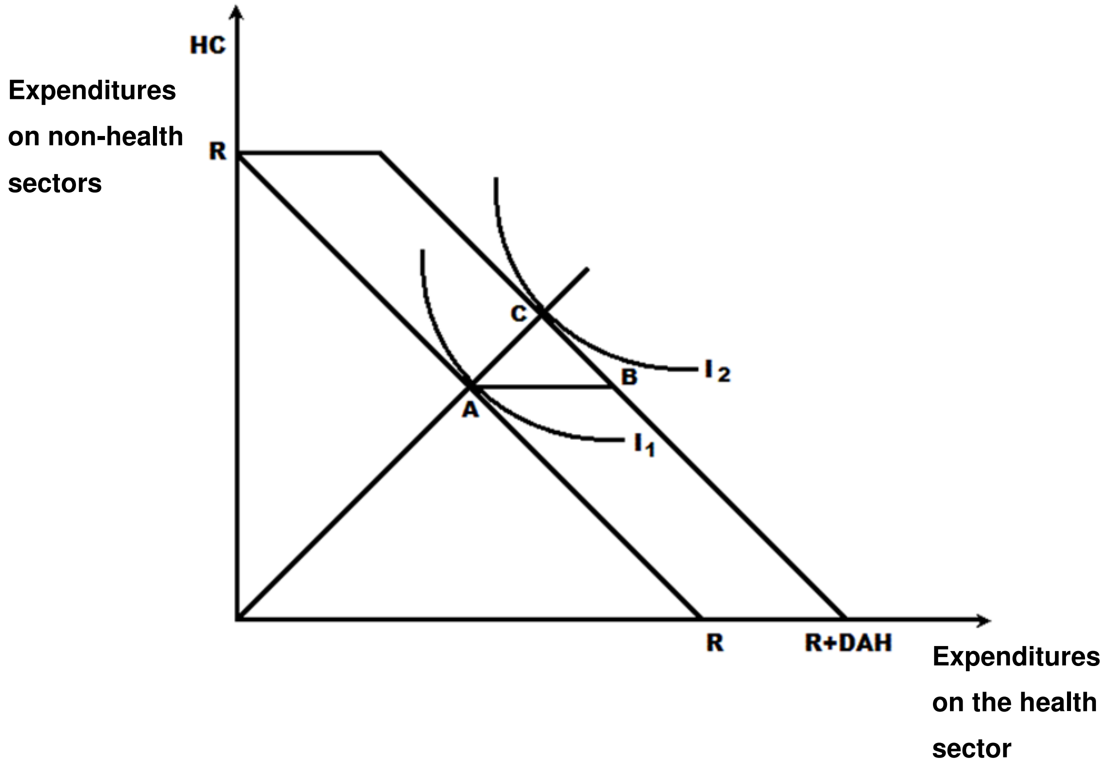
Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; year trends included, results suppressed for brevity.
[a] Baseline GDP/capita of no more than $1/day.
[b] Early SWAp implementation was defined as prior to 2003.
[c] Baseline control of corruption threshold set at -0.76 on the WGI control of corruption index, which ranges from about -2.5 to 2.5 (Kaufmann, Kraay et al. 2010). Note, the baseline observation for this indicator was actually 1996 (except for Solomon Islands, Cape Verde and Kiribati for whom it was 1998. A baseline score lower than this -0.76 threshold means that relatively poorer control of corruption is present in that country. This threshold was based upon the presence of an observed, obvious break in the data. An alternative threshold is applied in Supplementary Materials to assess the sensitivity of results to this decision (See Supplementary Materials, table D2).

Table 5. Evolving SWAp effects on displacement

	(1)	(2)
	GHE-S	GHE-S
$DAHG_{it}$	-0.810***	-0.833***
	(0.055)	(0.108)
$DAHNG_{it}$	0.101	0.059
	(0.098)	(0.107)
1-3 yrs pre_swap	0.003**	0.003**
	(0.001)	(0.001)
(Preswap)*dahg	0.350***	0.286*
	(0.110)	(0.153)
(Preswap)*dahng	0.000	0.000
	(.)	(.)
SWAp yrs 1-3	0.005***	0.006***
	(0.002)	(0.002)
(SWAp yrs 1-3)*dahg	0.433***	0.359*
	(0.157)	(0.210)
(SWAp yrs 1-3)*dahng	0.335**	0.375**
	(0.140)	(0.172)
SWAp yrs 4+	0.008***	0.009***
	(0.002)	(0.002)
(SWAp yrs 4+)*dahg	0.637***	0.585***
	(0.172)	(0.146)
(SWAp yrs 4+)*dahng	-0.211	-0.188
	(0.242)	(0.242)
Log $(GDP/capita)_{it}$	-0.003	-0.003
	(0.002)	(0.002)
GGE_{it}	0.028***	0.024**
	(0.010)	(0.012)
Log $(HIV\ deaths)_{it}$	0.000	0.000
	(0.001)	(0.001)
Log $(IMR)_{it}$		0.001
		(0.005)
Debt relief $_{it}$		0.001
		(0.003)
Government $effectiveness_{it}$		-0.000
		(0.002)
Corruption $control_{it}$		0.004**
		(0.002)
N	936	863
SWAp countries	27	27
Control countries	25	24

Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; year trends included, results suppressed for brevity.

Figure 1: Crowding out of government health expenditure



The impact of SWAps on aid displacement of domestic health expenditure - Appendix

Table 1. Sources of data

Variable	Description	Source
GHE-S	Government health expenditure as the source – funds drawn from government's own revenues (ie. Excluding DAH).	Constructed using WHO and IHME data
GHE-A	Government health expenditure as the agent – includes GHE-S and DAH channelled through government.	WHO
DAHG	Development assistance for health channelled to recipient government.	IHME
DAHNG	Development assistance for health channelled to non-government recipients in given country.	IHME
SWAp	Dummy variable indicating presence of SWAp.	Search of grey and peer reviewed literature. See web appendix for Sweeney, Mortimer et al. (2014)
GDP/capita	Gross domestic product per capita.	World Bank
GDP	Gross domestic product.	World Bank
GGE	General government's non-health expenditure.	Constructed using WHO data
HIV	HIV deaths per 100,000.	IHME
IMR	Infant mortality rate (deaths per 1,000 live births)	World Bank
Debt relief	Debt forgiveness or reduction.	World Bank
Government effectiveness	Index of perceived government effectiveness. Scores range between -2.5 (poorest) and 2.5 (best).	World Bank (see Kaufmann, Kraay et al. (2010))
Control of corruption	Index of perceived control corruption. Scores range between -2.5 (poorest) and 2.5 (best).	World Bank (see Kaufmann, Kraay et al. (2010))

References

Kaufmann, D., A. Kraay and M. Mastruzzi (2010). The worldwide governance indicators: methodology and analytical issues. Policy Research Working Paper No. 5430. World Bank: Development Research Group. Washington DC.

Sweeney, R., D. Mortimer and D. W. Johnston (2014). "Further investigations of the donor-flight response." Soc Sci Med **113**: 179-182.

Supplementary Materials:

The impact of SWAps on aid displacement of domestic health expenditure

A. Estimating propensity to SWap

Table A1. Linear probability model results for estimating propensity to ever implement a SWAp after 1996 given 1995 baseline observed characteristics.

	(1) SWAp
Log (GDP/capita)$_i$	-0.319**
	(0.126)
Log (GHE-A)$_i$	0.090
	(0.089)
Log (population)$_i$	-0.064
	(0.082)
Log (DAHG)$_i$	-0.014
	(0.011)
DAHNG$_i$	-0.052
	(0.067)
No. of donors$_i$	0.035*
	(0.019)
Log (IMR)$_i$	-0.010
	(0.146)
Life expectancy$_i$	0.012
	(0.012)
HIV deaths$_i$	0.000
	(0.001)
Africa – North of Sahara$_i$	-0.684**
	(0.319)
Sub-Saharan Africa$_i$	-0.303
	(0.221)
Europe$_i$	-0.674**
	(0.278)
Far East Asia$_i$	-0.444*
	(0.265)
Middle East$_i$	-0.713**
	(0.303)
North & Central America$_i$	-0.540**
	(0.235)
Oceania$_i$	-[a]
	-
South & Central Asia$_i$	-0.378
	(0.261)
South America$_i$	-0.575**
	(0.242)
Countries	99

[a] Oceania was the reference region. A significant negative coefficient for a Region suggests, holding other factors constant, an "average" country in that Region were less likely to implement a SWAp than had that "average" country been geographically located in Oceania.

B. ROBUSTNESS OF MAIN SPECIFICATION

Table B1. Models of health displacement without the moderating effects of SWAp[a].

	(1) Fixed Effects GHE-S	(6) Sys GMM[b] GHE-S
$GHE\text{-}S_{it-1}$	-	0.156**
		(0.073)
$DAHG_{it}$	-0.604***	-0.753***
	(0.109)	(0.102)
$DAHNG_{it}$	0.149	0.310***
	(0.111)	(0.120)
GGE_{it}	0.032***	0.055***
	(0.010)	(0.017)
Log $(GDP/capita)_{it}$	-0.003	0.002
	(0.002)	(0.002)
Log $(HIV\ deaths)_{it}$	-0.005	-0.000
	(0.007)	(0.001)
Observations	936	884
Countries	52	52
R^2	0.378	-
Instruments	-	51
Hansen *p-value*	-	0.204
AR(2) *p-value*	-	0.283

Standard errors in parentheses; * p < 0.10, ** p < 0.05, *** p < 0.01; year trends included, results suppressed for brevity.
[a] These models are comparable to existing literature and provide a comparative assessment of the external validity of the main specification. The estimates of DAHG displacement are smaller in magnitude but qualitatively consistent with Dieleman, Graves et al. (2013) estimates of displacement using fixed effects (-0.790) and System GMM (-0.856) estimation methods.
[b] Model used two-step estimation, and treats lagged GHE-S, DAHG and DAHNG as endogenous. The instrument set is collapsed and principal components analysis is employed to reduce the potential of instrument proliferation delivering a spurious Hansen test score.

<u>Sample Selection</u>

As can be seen in Table B2 below (which presents baseline characteristics of all health aid recipient countries meeting our data restrictions - described in table footnote), SWAps have been implemented in those countries who in 1995, were on average poorer and less healthy than the average non-implementing health aid recipient country. These differences motivated the sample selection approach set out in the manuscript. A comparison of baseline characteristics of the more inclusive sample described in Table B2 and the main sample (Table 2 of main text) shows this approach has identified a more closely matched set of countries (on observed factors at baseline). Further, Figure B1 compares the overlapping distribution of countries' propensity to SWAp scores (estimated via the linear probability model (LPM) set out in the main text) before (full sample of countries with sufficient data for inclusion in the LPM) and after sample selection on propensity scores. The footnote in Table B2 lists the inclusion criteria for the full sample. Table B3 provides the full list of countries included in the descriptive Table B2.

It is however important to test the sensitivity of main results to changes in sample selection decisions. Table B4 presents results from re-estimation of the main specification after applying the following list sample selection decisions (note sample number corresponds with column number in Table B4):

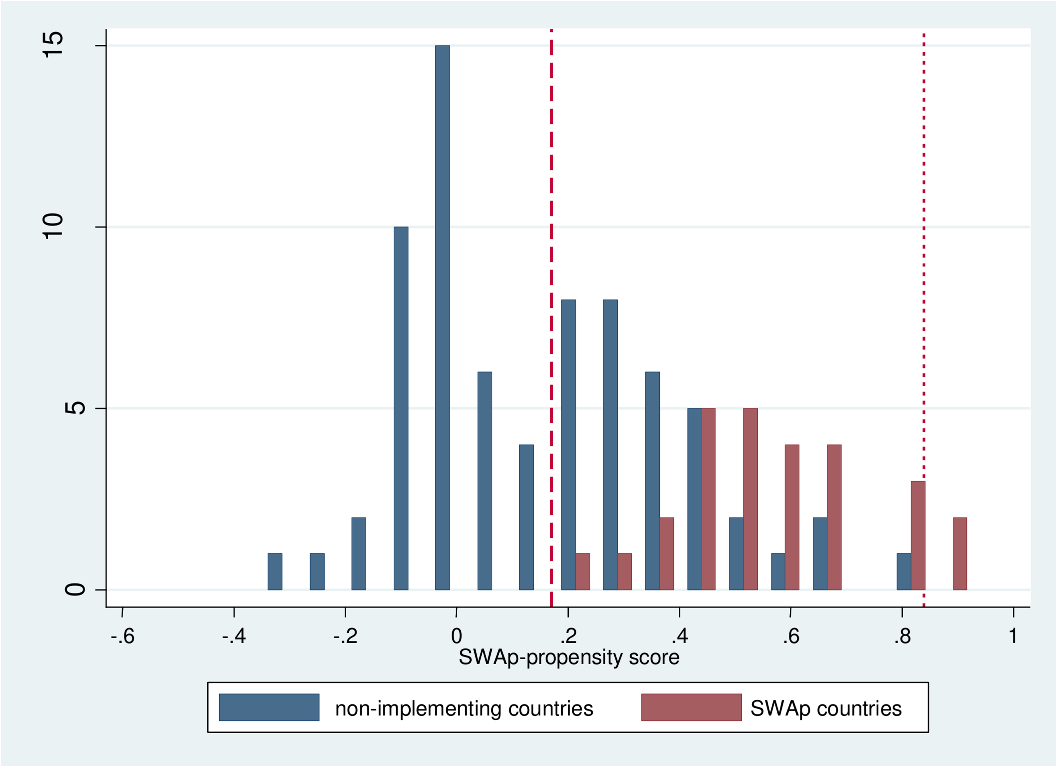
- Sample 1 = main sample results (i.e. all countries with SWAp-propensity score>0.24).
- Sample 2 = a more inclusive sample including countries with SWAp propensity score >0.2.
- Sample 3 = a more exclusive sample which restricts inclusion to SWAp-propensity scores from 0.24 > p > 0.86 (the overlapping range of SWAp propensity scores across groups).
- Sample 4 = most inclusive sample including all countries with sufficient data to include in the SWAp-propensity linear probability model (This sample is described in Table B2).
- Sample 5 = alternative sample selection approach based upon overlapping range of baseline GDP per capita - because health SWAps are most prominent in low-income health aid recipient countries (at baseline). Note baseline GDP per capita was measured as the average GDP per capita from 1990 to 1995.

Table B2. Baseline characteristics of larger sample of countries with "complete data"[a]

	SWAp Countries					Control Countries					
Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**	T-test p-value
GDP/capita ($)	27	411.71	301.49	134.40	1445.66	70	1771.17	1514.83	134.33	7407.65	0.00
Population (millions)	27	18.89	25.95	0.36	119.87	70	24.31	40.36	0.08	194.11	0.44
GHE-S/GDP (%)	27	1.33	1.42	-2.71	3.69	70	2.13	0.02	-0.08	0.09	0.03
DAHG/GDP (%)	27	0.83	0.95	0.00	3.74	70	0.55	0.02	0.00	0.11	0.28
DAHNG/GDP (%)	27	0.29	0.30	0.00	1.21	70	0.13	0.00	0.00	0.01	0.00
GGE/GDP ($)	27	20.47	8.51	5.51	52.25	70	26.33	0.14	0.07	0.77	0.01
IMR (per 1,000 live births)	27	84.39	29.57	29.10	143.40	70	52.87	35.17	8.40	153.40	0.00
Life expectancy	27	54.99	8.68	31.24	72.14	70	63.36	9.20	35.82	76.77	0.00
HIV deaths (per 100,000)	27	76.12	119.66	0.00	424.50	70	33.90	74.88	0.00	456.00	0.04

[a] Countries with at least 16 of a possible 18 complete annual observations (defined as GHE-S, DAHG, DAHNG, SWAp, GDP, and GDP/capita data present), including a 1995 (baseline). Further, countries required 1995 data on population, number of donors, infant mortality, life expectancy and HIV deaths for estimation of a SWAp-propensity score via the linear probability model. NOTE: Zambia excluded as its SWAp commenced prior to first GHE-S observation. India and China were excluded given country-wide SWAp very unlikely. Lithuania, Latvia and Poland were included in the LPM given they had received DAHNG in the period of interest and thus may influence SWAp likelihood, but were excluded from subsequent analyses as had received zero DAHG disbursements.

Figure B1. Comparison of SWAp-propensity score distributions by SWAp status



-

Note: All countries with propensity score to the right of the far left vertical (dashed) line are included in the main sample. The more exclusive Sample 3 includes all countries that fall within the range of the two vertical (dashed and dotted) lines. Sample 4 – the most inclusive sample includes all countries with sufficient data to generate a SWAp propensity score (i.e. all countries in Figure B1).

Table B3. Countries used in selected alternative samples to test sensitivity main findings to sample selection decisions

SWAp countries (n=27) [a b]		**Non-implementing countries[a](n=70)**			
Bangladesh	Mozambique	Albania[b]	Congo[b]	Jamaica	Russia
Benin	Nepal	Angola[b]	Congo, DRC[b]	Jordan	Sierra Leone[b]
Burkina Faso	Nicaragua	Argentina	Costa Rica	Kazakhstan	South Africa
Burundi	Niger	Armenia[b]	Cote d'Ivoire[b]	Kenya[b]	Sri Lanka[b]
Cambodia	Papua New	Belize	Cuba	Kiribati[b]	St.Vincent &
Ethiopia	Guinea	Bhutan[b]	Dominican	Laos[b]	the
Ghana	Rwanda	Bolivia[b]	Republic	Lebanon	Grenadines
Kyrgyzstan	Senegal	Bosnia &	Ecuador	Malaysia	Suriname[b]
Lesotho	Solomon Is.		Egypt[b]	Mauritius	Swaziland
Madagascar	Sudan	Herzegovina[b]	El Salvador	Mexico	Thailand
Malawi	Tanzania[c]	Botswana	Eritrea	Morocco	The Gambia[b]
Mali	Uganda	Brazil	Fiji	Namibia	Togo[b]
Mauritania	Uzbekistan	Cameroon[b]	Guatemala[b]	Nigeria[b]	Tunisia
Mongolia	Vietnam	Cape Verde[b]	Guinea[b]	Pakistan[b]	Turkey
		Central African	Guinea-Bissau[b]	Panama	Uruguay
		Republic[b]	Guyana[b]	Paraguay	Venezuela
		Chad[b]	Haiti[b]	Peru	Yemen[b]
		Chile	Honduras[b]	Philippines[b]	
		Colombia	Indonesia[b]	Romania	
		Comoros[b]	Iran		

[a] Countries meeting "complete data" requirements set out in footnote of Table B2.

[b] These countries are included in an alternative robustness sample (Sample 5 above), selected on the basis of the overlapping range (of SWAp and non-implementing countries) of baseline GDP per capita. Baseline GDP per capita was measured as the average GDP per capita from 1990 to 1995. This selection process also identified the following countries for inclusion that had insufficient data to be included in the linear probability model sample selection approach: SWAp countries – Moldova and Samoa; Non-SWAp countries – Azerbaijan, Djibouti, Georgia and Turkmenistan.

[c] In an additional test of the sensitivity of main specification results to sample selection the inclusion SWAp-propensity score thresholds were set at 0.24 > p > 0.86, resulting in Tanzania's exclusion with a SWAp-propensity score >0.86 (Sample 3 above).

Table B4. Sensitivity of findings to alternative sample selection decisions

	(1) Main Sample Propensity>0.24 GHE-S	(2) Sample 2 Propensity>0.2 GHE-S	(3) Sample 3 Propensity range: 0.24 to 0.86 GHE-S	(4) Sample 4 "complete" data countries[a] GHE-S	(4) Sample 5 Baseline GDP/capita: \$156 to \$1184[b] GHE-S
$DAHG_{it}$	-0.779***	-0.779***	-0.779***	-0.752***	-0.788***
	(0.061)	(0.061)	(0.061)	(0.065)	(0.061)
$SWAp_{it}$	0.005***	0.004***	0.005***	0.002	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$SWAp*DAHG_{it}$	0.518***	0.525***	0.530***	0.519***	0.520***
	(0.123)	(0.122)	(0.125)	(0.122)	(0.122)
$DAHNG_{it}$	0.124	0.137	0.127	0.083	0.107
	(0.103)	(0.102)	(0.103)	(0.096)	(0.112)
$SWAP*DAHNG_{it}$	-0.084	-0.092	-0.127	-0.084	-0.076
	(0.198)	(0.196)	(0.203)	(0.193)	(0.200)
Log (GDP/capita)$_{it}$	-0.003	-0.002	-0.003	-0.000	-0.002
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)
GGE_{it}	0.029***	0.030***	0.028***	0.041***	0.041***
	(0.010)	(0.009)	(0.010)	(0.010)	(0.010)
Log (HIV deaths)$_{it}$	0.000	-0.000	0.000	-0.001	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	936	1044	918	1742	1206
SWAp countries	27	27	26	27	29
Control countries	25	31	25	70	38

Standard errors in parentheses; * p < 0.10, ** p < 0.05, *** p < 0.01; year trends included, results suppressed for brevity.

[a] most inclusive sample including all countries with sufficient data to include in the SWAp-propensity linear probability model.

[b] These countries are included in an alternative robustness sample (Sample 5 above), selected on the basis of the overlapping range (of SWAp and non-implementing countries) of baseline GDP per capita. Baseline GDP per capita was measured as the average GDP per capita from 1990 to 1995

Table B5. Dynamic panel estimation of SWAp*DAH effect using SysGMM

	OLS	FE	SYS-GMM[a]	
	(1)	(2)	(3)[bc]	(4)[d]
	GHE-S	GHE-S	GHE-S	GHE-S
GHE-S$_{t-1}$	0.652***	0.249***	0.218**	0.235***
	(0.079)	(0.073)	(0.087)	(0.072)
DAHG$_{it}$	-0.426***	-0.714***	-0.620***	-0.679***
	(0.133)	(0.083)	(0.147)	(0.090)
SWAp$_{it}$	0.003***	0.004***	0.004**	0.003**
	(0.001)	(0.001)	(0.002)	(0.001)
SWAp*DAHG$_{it}$	0.320***	0.420***	0.300**	0.659***
	(0.110)	(0.106)	(0.142)	(0.179)
DAHNG$_{it}$	0.330***	0.151	0.509***	0.562***
	(0.121)	(0.100)	(0.147)	(0.126)
SWAP*DAHNG$_{it}$	-0.291**	-0.087	-0.331	-0.597**
	(0.138)	(0.171)	(0.269)	(0.255)
Log (GDP/capita)$_{it}$	0.001	-0.002	0.002	0.002*
	(0.001)	(0.002)	(0.002)	(0.001)
GGE$_{it}$	0.024**	0.023**	0.049***	0.047***
	(0.010)	(0.009)	(0.017)	(0.015)
Log (HIV deaths)$_{it}$	-0.000	0.000	-0.000	-0.000
	(0.000)	(0.001)	(0.001)	(0.001)
Observations	884	884	884	884
SWAp countries	27	27	27	27
Control countries	25	25	25	25
Instruments	-	-	58	74
Hansen *p-value*	-	-	0.239	0.961
AR(2) *p-value*	-	-	0.262	0.176

Windmeijer standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; year trends included, results suppressed for brevity.

[a] All models employ two-step estimation, collapsed instrument sets and principal components analysis to further reduce instrument numbers.

[b] Preferred model based upon Roodman's (2009a) guidelines for model selection:

- Estimate of lagged dependent variable falls within (or close to) coefficient estimates for OLS and fixed effects dynamic models presented in columns 1 and 2.
- Failure to reject Hansen test of over-identification restrictions with p-value that falls within (or near to) desired range between 0·1 and 0·25. Note, as Hansen p values get closer to 1·00, suggests too many instruments are over-fitting the model.
- Model should accept the null that there is no 2nd order autocorrelation.

[c] Lagged GHE-S, DAHG, DAHNG treated as endogenous.

[d] Lagged GHE-S, DAHG, DAHNG, SWAp*DAHG, SWAP*DAHNG treated as endogenous.

Measurement Error

As discussed in the Manuscript, Errors in Variables (EIV) models have been run to assess the sensitivity of the main findings to an alternative modelling approach designed to remove the bias associated with the presence of potential measurement in DAHG, DAHNG and consequently DAHG*SWAp and DAHNG*SWAp. We explore this sensitivity running additional regressions using the xtewreg Stata commands. The xtewreg command employs a two-step GMM estimation of the EIV model using higher-order moments/cumulants of residuals to 'correct' for errors in one or more of the RHS variables, (Erickson & Whited, 2002; Erickson & Whited, 2012; Erickson et al, forthcoming). We use GHE-A as the dependent variable to exclude the possibility of mismeasurement on the left hand side arising as a consequence of mismeasurement in DAHG. As such, the interpretation of the DAHG coefficient is slightly different. When we regress DAHG on GHE-A, we would expect a coefficient of 1 (rather than zero for GHE-S) in the event of full additionality of DAHG. Because DAHNG is not reflected in government health expenditure, we would expect *coefficients of* 0 for regressions on both GHE-S and GHE-A in the event of full additionality for DAHNG.

This EIV modelling technique is known to be less stable in small samples – so we conduct this analysis in the larger "Sample 4" established above to maximise likelihood of model stability. In Table B6, we first run a series of EIV regressions in a variation of the main specification that excludes the mediating effect of SWAp on crowding out (by excluding the SWAp and SWAp-interaction terms). We assume both DAHG and DAHNG are mismeasured. In Table B7 we estimate the main effect of interest incorporating the SWAp and SWAp interaction terms, where we assume DAHG, SWAp*DAHG, DAHNG and SWAp*DAHNG are mismeasured. xtewreg does not compute fixed effects internally and so, for each set of EIV models, we undertake two alternative approaches for controlling for country-specific time-invariant factors. First all data is modelled as variations from country means (i.e. we demean all data using country-specific means). Second, we use the raw data but include country dummy terms. In each table we also present in the first three columns results from variants of our main model: (1) fixed effects model using GHE-S as the dependent variable (the main model), (2) fixed effects model using GHE-A as the dependent variable (main model with different dependent variable), and (3) OLS on demeaned data using GHE-A as the dependant variable. In Table B8 we extend the EIV model analyses, employing the eivreg Stata command. This approach requires a predetermined level of "reliability" of assumed mismeasured variables (DAHG, DAHNG, SWAp*DAHG and SWAp*DAHNG). Given levels of reliability are contested and unknown, we test sensitivity of main results to differing (but uniform) levels of reliability across mismeasured variables. Specifically reliability levels are modelled first at 0.9 and 0.8 (models with lower levels of estimated reliability than this are not estimated by eivreg as it does not estimate models where the reliability is less than the R^2).

In Table B6 only one model – Column 8, meets guidelines for model selection (i.e.(a) tau^2 scores for each mismeasured variable's proxy falling close to, but not above 1 (proxy tau^2 scores should fall between 0 = worthless proxy and 1 = perfect proxy); and (b) non-significant Sargan/Hansen p-score, which fails to reject the null that the model is overidentified). In this model, the extent of displacement of government health expenditure remains consistent with a crowding out effect of domestically sourced government health expenditure (coefficient is significantly different to 1, p=0.00). As can be seen in Table B7, no EIV model satisfied criteria for model selection – including some mismeasured variable's proxies tau^2 scores falling outside the plausible range of 0 and 1, so any conclusions drawn should be treated with caution. While the magnitude of the SWAp*DAHG effect is sensitive to EIV model selection, it is statistically significant in most models and remains qualitatively consistent with results from the main model. Further, Table B8 showed the main SWAp*DAHG effect was robust to the differing levels of assumed reliability of mismeasured variables. Nonetheless, the extent of sensitivity around important coefficients in both Tables B6 and B7 – namely DAHG and DAHNG – suggest results should be treated with some caution.[1] Given this, it is difficult to conclude that these models have satisfactorily controlled for the presence of mismeasurement. Further research is needed to investigate the finite sample properties of these estimators when using small N samples to estimate relatively complex EIV models in which a number of RHS variables are potentially mismeasured.[2] Whilst on the basis of this robustness analyses, we cannot conclude with full certainty that

[1] This is not surprising when using small N samples to estimate relatively complex EIV models in which a number of RHS variables are potentially mismeasured. Precise estimation of higher-order moments/cumulants requires relatively large N and data requirements increase for each additional RHS variable specified as mis-measured. Moreover, parameter estimates from this approach are potentially sensitive to mis-specification of RHS variables as either mis-measured or perfectly measured (Erickson et al, forthcoming).

[2] While the Erickson & Whited two-step GMM models perform well when estimating relatively simple specifications with one mismeasured RHS variable in large N samples (Erickson & Whited, 2012), the models estimated here include several RHS variables with potentially significant measurement error and our data spans a relatively small N sample over the relatively recent period in which (relatively) good quality data have become available. Erickson & Whited (2012) note that "Monte Carlo experiments are only informative about an estimator at one point in a parameter space and for one joint distribution of the data" and recommend "examining finite sample performance in data-relevant Monte Carlos before using them on new data, especially if the sample size differs greatly from the ones we consider here". And so, while our results from Erickson & Whited EIV models demonstrate the potential importance of measurement error and the potential value of this approach, further work is required before drawing strong conclusions from the sometimes conflicting results that arise from these models.

SWAp has significantly reduced the extent of crowding out, nor can we exclude this effect. The extent of mismeasurement remains contested and the data used is argued to have made significant gains in accuracy since Van de Sijpe and others first raised these important measurement concerns. As such, there appears to be sufficient evidence here to motivate further investigation of the potential for SWAp to address the (perceived) threat of DAHG displacement of domestically sourced government health expenditure.

Table B6. Sensitivity of crowding out models excluding the mediating effects of SWAp using alternative "Errors in Variables" modelling approach

	OLS models			EIV models – demeaned data			EIV models – country FE models		
	(1)	(2)	(3)	(4)[a]	(5)[b]	(6)[c]	(7)[a]	(8)[b]	(9)[c]
	FE	FE	Demeaned	EIV 3(2)	EIV 4(2)	EIV 5(2)	EIV 3(2)	EIV 4(2)	EIV 5(2)
	GHE-S	GHE-A	GHE-A	GHE-A	GHE-A	GHE-A	GHE-A	GHE-A	GHE-A
$DAHG_{it}$	-0.591***	0.409***	0.336***	0.828***	0.157***	0.121***	-0.073	0. 441***	0.120***
	(0.104)	(0.104)	(0.100)	(0.283)	(0.031)	(0.008)	(0.334)	(0.072)	(0.017)
$DAHNG_{it}$	0.153	0.153	0.168	-0.963	1.192***	-0.267***	2.362	0.973***	1.548***
	(0.115)	(0.115)	(0.113)	(1.319)	(0.178)	(0.002)	(2.597)	(0.116)	(0.068)
Log (GDP/capita)$_{it}$	-0.000	-0.000	-0.000	-0.001	0.002	-0.002	0.007*	0.006***	0.006***
	(0.002)	(0.002)	(0.001)	(0.004)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)
GGE_{it}	0.041***	0.041***	0.042***	0.045***	0.037***	0.046***	0.035*	0.040***	0.038***
	(0.010)	(0.010)	(0.010)	(0.013)	(0.011)	(0.010)	(0.019)	(0.011)	(0.011)
Log (HIV deaths)$_{it}$	-0.001	-0.001	-0.001	-0.001	0.000	-0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	1742	1742	1742	1742	1742	1742	1742	1742	1742
SWAp countries	27	27	27	27	27	27	27	27	27
Control countries	70	70	70	70	70	70	70	70	70
$tau1^{2}$[d]				0.515(0.275)	1.730(0.690)	2.840(0.848)	-0.737(7.100)	0.943(0.116)	1.668(0.435)
$tau2^{2}$				0.056 (0.173)	0.282(0.108)	-0.491(0.433)	0.644(0.109)	0.681(0.087)	0.663(0.082)
Sargan/Hansen[e] p-score=				0.792	0.056	0.001	0.203	0.249	0.001

Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; year trends included, results suppressed for brevity.

[a] 3rd order moment and assumes DAHG and DAHNG are mis-measured.

[b] 4th order moment and assumes DAHG and DAHNG are mis-measured.

[c] 5th order moment and assumes DAHG and DAHNG are mis-measured.

[d] Tau^2 scores should fall within the range 0 and 1, where 0 indicates a worthless proxy and 1 indicates a perfect proxy. Tau^2 scores outside that range may indicate the sample size is too small.

[e] Sargan Hansen test of overidentifying restrictions of the model. The null is that the model is not overidentified.

Table B7. Sensitivity of main results to alternative "Errors in Variables" modelling approach using Stata command xtewreg

	OLS models			EIV models – demeaned data		EIV models – country FE models	
	(1)	(2)	(3)	(4)[a]	(5)[b]	(6)[c]	(7)[d]
	FE	FE	Demeaned	EIV 3(4)	EIV 4(4)	EIV 3(4)	EIV 4(4)
	GHE-S	GHE-A	GHE-A	GHE-A	GHE-A	GHE-A	GHE-A
$DAHG_{it}$	-0.752***	0.248***	0.230***	-0.064	0.287***	0.074	0. 708***
	(0.065)	(0.065)	(0.067)	(0.295)	(0.013)	(0.097)	(0.022)
$SWAp_{it}$	0.002	0.002	0.004***	0.004**	0.004***	0.003	0.001
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)
$SWAp*DAHG_{it}$	0.519***	0.519***	0.448***	0.381	0.720***	0.236***	0.188***
	(0.122)	(0.122)	(0.157)	(0.294)	(0.020)	(0.098)	(0.019)
$DAHNG_{it}$	0.083	0.083	0.108	-0.008	0.269***	0.508***	-1.238**
	(0.096)	(0.096)	(0.101)	(0.122)	(0.026)	(0.106)	(0.063)
$SWAP*DAHNG_{it}$	-0.084	-0.084	-0.027	0.650***	-0.429***	0.135	0.997***
	(0.193)	(0.193)	(0.170)	(0.141)	(0.021)	(0.129)	(0. 047)
Log $(GDP/capita)_{it}$	-0.000	-0.000	-0.000	-0.002	0.000	0.005*	0.004
	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.003)	(0.003)
GGE_{it}	0.041***	0.041***	0.041***	0.042***	0.040***	0.043***	0.048***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.012)
Log $(HIV\ deaths)_{it}$	-0.001	-0.001	-0.001*	-0.001	-0.001	0.001	0.001
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Observations	1742	1742	1742	1742	1742	1742	1742
SWAp countries	27	27	27	27	27	27	27
Control countries	70	70	70	70	70	70	70
tau1[2d]				-3.094(15.24)	0.674(0.164)	2.248(2.184)	0.836(0.082)
$tau2^2$				6.624(93.29)	0.786(0.28)	1.553(0.475)	0.947(0.214)
$tau3^2$				1.494(0.7)	0.665(0.18)	0.720(0.128)	0.676(0.074)
$tau4^2$				0.367(0.33)	0.716(0.24)	0.240(1.134)	1.057(0.064)
Sargan/Hansen[e] p-score=				0.435	0.000	0.599	0.000

Standard errors in parentheses; * p < 0.10, ** p < 0.05, *** p < 0.01; year trends included, results suppressed for brevity.

[a] 3rd order moment and assumes DAHG, DAHNG, SWAp*DAHG and SWAp*DAHNG are mis-measured.

[b] 4th order moment and assumes DAHG, DAHNG, SWAp*DAHG and SWAp*DAHNG.

[c] 5th order moment and assumes DAHG, DAHNG, SWAp*DAHG and SWAp*DAHNG.

[d] Tau^2 scores should fall within the range 0 and 1, where 0 indicates a worthless proxy and 1 indicates a perfect proxy. Tau^2 scores outside that range may indicate the sample size is too small.

[e] Sargan Hansen test of overidentifying restrictions of the model. The null is that the model is not overidentified.

Table B8. Sensitivity of main results to alternative "Errors in Variables" modelling approach using Stata command eivreg

	Models with no SWAp or SWAp interactions				Models with SWAp and SWAp interactions			
	Demeaned data models[a]		Country fixed effects models[b]		Demeaned data models[a]		Country fixed effects models[b]	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	r^c=0.9	r^c=0.8	r^c=0.9	r^c=0.8	r^d=0.9	r^d=0.8	r^d=0.9	r^d=0.8
	GHE-A	GHE-A	GHE-A	GHE-A	GHE-A	GHE-A	GHE-A	GHE-A
$DAHG_{it}$	0.375***	0.425***	0.591***	0.906	0.265***	0.313***	0.402***	0.762***
	(0.030)	(0.034)	(0.042)	(0.063)	(0.033)	(0.039)	(0.049)	(0.098)
$SWAp_{it}$	-	-	-	-	0.004***	0.004***	0.002*	0.002***
					(0.001)	(0.001)	(0.001)	(0.001)
$SWAp*DAHG_{it}$	-	-	-	-	0.415***	0.370***	0.461***	0.216**
					(0.067)	(0.070)	(0.064)	(0.092)
$DAHNG_{it}$	0.183***	0.202***	0.250***	0.251***	0.128**	0.156**	0.200**	0.284
	(0.048)	(0.055)	(0.060)	(0.089)	(0.059)	(0.074)	(0.081)	(0.179)
$SWAP*DAHNG_{it}$	-	-	-	-	-0.043	-0.067	-0.116	-0.143
					(0.092)	(0.100)	(0.104)	(0.177)
Log (GDP/capita)$_{it}$	0.000	0.000	0.005***	0.006***	-0.000	0.000	0.005***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
GGE_{it}	0.042***	0.041***	0.043***	0.040***	0.040***	0.040***	0.043***	0.040***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Log (HIV deaths)$_{it}$	-0.001**	-0.001**	0.001***	0.001***	-0.001**	-0.001**	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	1742	1742	1742	1742	1742	1742	1742	1742
SWAp countries	27	27	27	27	27	27	27	27
Control countries	70	70	70	70	70	70	70	70

Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; year trends included, results suppressed for brevity.

[a] Country time invariant factors controlled by demeaning data with country specific means.

[b] Country time invariant factors controlled by inclusion of country dummies.

[c] Estimated reliability of assumed mismeasured variables (DAHG and DAHNG)

[d] Estimated reliability of assumed mismeasured variables (DAHG, DAHNG, SWAp*DAHG and SWAp*DAHNG)

Table B9. Sensitivity of results to alternative HIV indicators (in log) (including data source)

	(1) HIV deaths (IHME) GHE-S	(2) HIV prevalence (World Bank) GHE-S	(3) HIV prevalence[a] (IHME) GHE-S	(4) HIV prevalence[a] (IHME) GHE-S
$DAHG_{it}$	-0.779***	-0.801***	-0.787***	-0.779***
	(0.061)	(0.066)	(0.056)	(0.061)
$SWAp_{it}$	0.005***	0.005***	0.005**	0.005***
	(0.001)	(0.001)	(0.002)	(0.001)
$SWAp*DAHG_{it}$	0.518***	0.593***	0.690**	0.518***
	(0.123)	(0.127)	(0.326)	(0.122)
DAHNG	0.124	0.119	0.089	0.123
	(0.103)	(0.126)	(0.140)	(0.102)
SWAP*DAHNG	-0.084	-0.147	-0.083	-0.084
	(0.198)	(0.212)	(0.365)	(0.198)
Log (GDP/capita) $_{it}$	-0.003	-0.002	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)
GGE_{it}	0.029***	0.031***	0.038***	0.029***
	(0.010)	(0.011)	(0.011)	(0.010)
Log (HIV burden) $_{it}$	0.000	-0.001	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.000)
N	936	846	208	935
SWAp countries	27	25	27	27
Control countries	25	22	25	25

Standard errors in parentheses; * p < 0.10, ** p < 0.05, *** p < 0.01; year trends included, results suppressed for brevity.
[a] IHME HIV prevalence data estimates available include observations in 1995, 2000, 2005, 2010 and 2013. Column 4 imputes observations for these missing values by interpolating linear trends between observed years.

FIGURE B2. Sensitivity to variations of SWAp commencement and exclusion of each SWAp

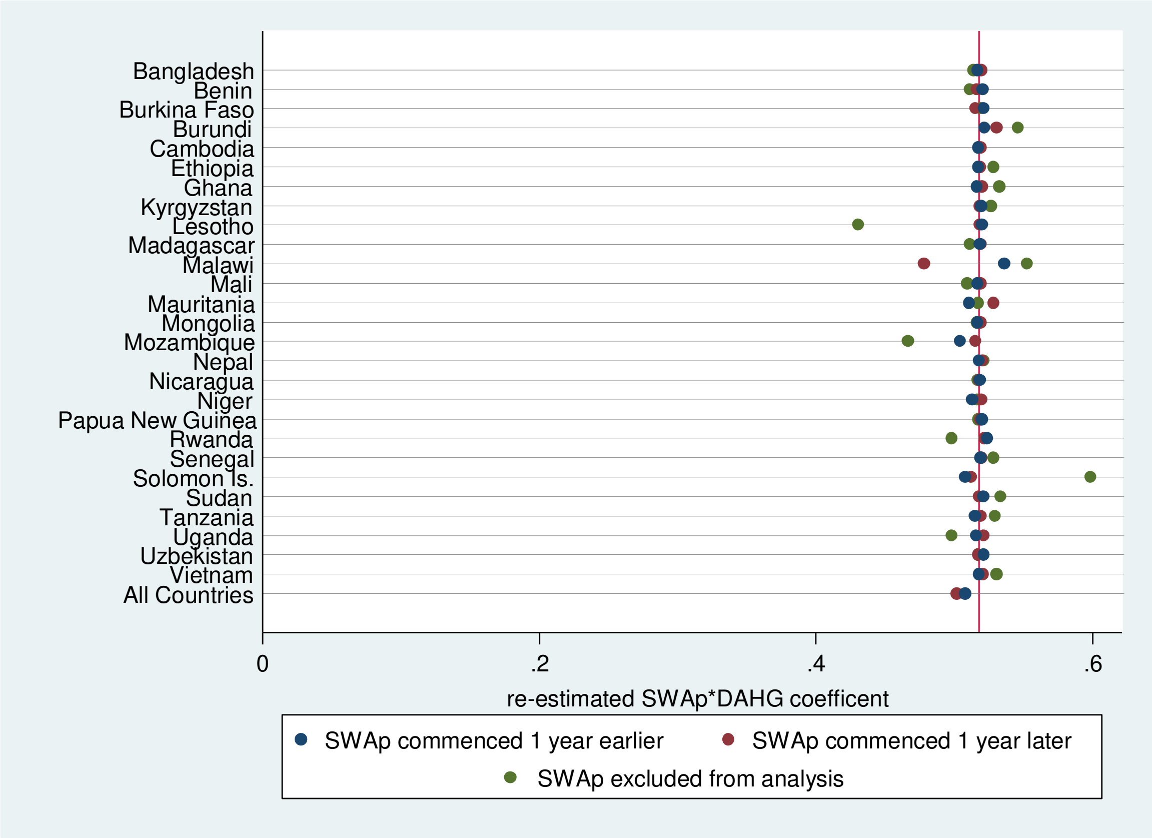


Figure B2 plots the re-estimated coefficients for the SWAp interaction with DAHG. As can be seen the estimated protective effect of SWAp presented in the paper (reducing the extent of displacement by $0.52) is robust to plausible variation in each SWAps commencement, including commencing all SWAps one year earlier and one year later. The finding appears most sensitive to the exclusion of the Solomon Islands SWAp and of more concern given the subsequent reduced effect size, the exclusion of the Lesotho and Mozambique SWAps. However, the replacement effect of increasing DAHG via SWAp when these countries were individually and then both excluded remained significant, all with p values less than 0.01. This analysis gives confidence that the significant effect of SWAp on the displacement effect of DAHG is robust to plausible variation in SWAp commencement uncertainty and not driven by the misclassification of any individual SWAp country.

C. Testing the common trends assumption

To test adherence to the common trends assumption requires – at a minimum – ability to compare GHE-S trends in the treatment (SWAp implementing) and control group over a lead up period, prior to any included SWAp uptake. Unfortunately, the WHO GHE-A data used in the main analyses commences in 1995, whilst the first SWAps included in this analysis commenced in 1997 when "preparation to SWAp responses" may already have become observable (as it appears in the evolving SWAp effect analyses). To address this limitation and thereby enable some comparisons of pre-Swap uptake GHE-S trends we employed an alternative public health expenditure data from the International Food Policy Research Institute and constructed a panel from 1990 to 1996 for analysis (IFPRI 2015). A limitation of this approach is that sufficient data was only available for 24 of our main sample countries (both groups had a number of countries represented). Nonetheless, it does enable some assessment of the presence of pre-existing differences in trends between the treatment (SWAp countries) and control groups prior to any included SWAp implementation. With this alternative dataset, two tests were undertaken to assess adherence to the common trends assumption. An amended version of Equation (1) was estimated, however the SWAp and SWAp*DAH interaction terms were replaced by (i) SWAp*time interaction to identify the time-trend in the SWAp group; and (ii) a SWAp dummy that introduced a fake SWAp to all SWAp group countries in 1993. In both tests, support for adherence to the common trends assumption requires no significant effect associated with the respective SWAp identifier variable.

As can be seen in Table C1 below, the magnitude of the displacement effect of DAHG on GHE-S in the pre-SWAp period 1990 to 1996 was larger than that estimated by this present paper and the more recent Dieleman et al (2013) paper, though the estimate is consistent with upper limits of Lu, Schneider et al. (2010) and thus may reflect the limitations of some of these older data point estimates as well as being influenced by the smaller set of countries. Notwithstanding the data limitations, the DAHG displacement estimates are qualitatively consistent with estimates from the observed period of 1995 to 2012 presented in the paper. Most importantly, there was no observed difference in the pre-SWAp uptake GHE-S trends between the treatment and control groups (column 1, Table C1) nor did the fake SWAp identify any effective difference between the two (column 2, Table C1). So, whilst data limitations mean uncertainties remain, these tests provide support that the common trends assumption holds.

TABLE C1. Tests of the GHE-S trends of SWAp and control group between 1990 to 1996 (prior to SWAp uptake)[a]

	(1) GHE-S[b]	(2) GHE-S[b]
$DAHG_{it}$	-0.947***	-0.946***
	(0.034)	(0.035)
SWAp-group (linear trend)$_{it}$	-0.000	-
	(0.000)	-
Fake-$SWAp_{it}$	-	-0.001
	-	(0.001)
$DAHNG_{it}$	-0.414*	-0.418*
	(0.203)	(0.206)
Log (GDP/capita)$_{it}$	-0.000	-0.000
	(0.002)	(0.002)
Log (HIV deaths)$_{it}$	-0.002*	-0.002*
	(0.001)	(0.001)
N	159	159
SWAp countries	15	15
Control countries	9	9

Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Year trends included, results suppressed for brevity.

[a] Note: general government non-health expenditure has been excluded from this specification as it also had no observations prior to 1995.

[b] It was uncertain from accompanying notes if IFPRI (2015) estimates of public health expenditure included DAHG (ie. Whether it was GHE-S or GHE-A). Re-running the main model (from1995 to 2012) with the IFPRI GHE data with and without DAHG first subtracted indicated that their estimates likely included DAHG (GHE-A). Thus the results presented in Table C1 include the regression in which DAHG has first been subtracted from GHE data (to measure GHE-S).

D. Robustness of poorest subgroup analysis

Table D1. Robustness of poorest subgroup analysis to alternative POOREST sample

	(1) Poorest Sample GDP/capita ≤ \$365[ab] GHE-S	(2) Sample 2 GDP/capita ≤ \$456[b] GHE-S
$DAHG_{it}$	-0.860***	-0.855***
	(0.063)	(0.062)
$SWAp_{it}$	0.006***	0.006***
	(0.002)	(0.002)
$SWAp*DAHG_{it}$	0.543***	0.517***
	(0.093)	(0.090)
$DAHNG_{it}$	0.127	0.102
	(0.155)	(0.154)
$SWAP*DAHNG_{it}$	-0.143	-0.073
	(0.232)	(0.227)
Log (GDP/capita)$_{it}$	-0.007**	-0.003
	(0.003)	(0.003)
GGE_{it}	0.028**	0.025*
	(0.012)	(0.013)
Log (HIV deaths)$_{it}$	-0.001	-0.001
	(0.002)	(0.001)
N	504	558
SWAp countries	17	18
Control countries	11	13

Standard errors in parentheses; * p < 0.10, ** p < 0.05, *** p < 0.01; year trends included, results suppressed for brevity.

[a] Result presented in paper.

[b] at baseline.

Table D2. Robustness of subgroup analysis of underlying corruption to alternative threshold.

	Baseline control of corruption threshold score of -0.76		Baseline control of corruption threshold score of -0.96[a]	
	(1) Poorer control of corruption GHE-S	(2) Stronger control of corruption GHE-S	(3) Poorer control of corruption GHE-S	(4) Stronger control of corruption GHE-S
$DAHG_{it}$	-0.798***	-0.762***	-0.807***	-0.765***
	(0.064)	(0.105)	(0.072)	(0.094)
$SWAp_{it}$	0.005**	0.004***	0.003*	0.005***
	(0.002)	(0.001)	(0.002)	(0.001)
$SWAp*DAHG_{it}$	0.253	0.544***	0.038	0.565***
	(0.210)	(0.140)	(0.156)	(0.137)
$DAHNG_{it}$	0.135	0.141	-0.099	0.190**
	(0.205)	(0.086)	(0.237)	(0.092)
$SWAP*DAHNG_{it}$	0.408*	-0.308	0.516*	-0.220
	(0.213)	(0.215)	(0.264)	(0.219)
$\text{Log (GDP/capita)}_{it}$	-0.002	-0.003	0.000	-0.003
	(0.003)	(0.002)	(0.003)	(0.002)
GGE_{it}	0.032*	0.024***	0.044***	0.023***
	(0.016)	(0.006)	(0.014)	(0.007)
$\text{Log (HIV deaths)}_{it}$	-0.001	0.001	-0.001	0.000
	(0.002)	(0.001)	(0.002)	(0.001)
N	396	540	306	630
SWAp countries	9	18	7	20
Control countries	13	12	10	15

Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; year trends included, results suppressed for brevity.

[a] Alternative threshold reflected an obvious break in baseline corruption index scores in the SWAp group.

E. References

Dieleman, J. L., C. M. Graves and M. Hanlon (2013). "The Fungibility of Health Aid: Reconsidering the Reconsidered." The Journal of Development Studies **49**(12): 1755-1762.

International Food Policy Research Institute (IFPRI). 2015. Statistics on public expenditures and economic development (SPEED). Washington, D.C.: International Food Policy Research Institute. http://dx.doi.org/10.7910/DVN/INZ3QK

Kaufmann, D., A. Kraay and M. Mastruzzi (2010). The worldwide governance indicators: methodology and analytical issues. Policy Research Working Paper No. 5430. World Bank: Development Research Group. Washington DC.

Lu, C., M. T. Schneider, P. Gubbins, K. Leach-Kemon, D. Jamison and C. J. L. Murray (2010). "Public financing of health in developing countries: a cross-national systematic analysis." The Lancet **375**(9723): 1375-1387.

Roodman, D. (2009a). "How to do xtabond2: An introduction to difference and system GMM in Stata." Stata Journal **9**(1): 86-136.

Sweeney, R., D. Mortimer and D. W. Johnston (2014). "Further investigations of the donor-flight response." Soc Sci Med **113**: 179-182.